Популярные лекции
ПО МАТЕМАТИКЕ

Ю. А. ШРЕЙДЕР

ЧТО ТАКОЕ РАССТОЯНИЕ?

ПОПУЛЯРНЫЕ ЛЕКЦИИ ПО МАТЕМАТИКЕ
ВЫПУСК 38

Ю. А. ШРЕЙДЕР

ЧТО ТАКОЕ РАССТОЯНИЕ?

ГОСУДАРСТВЕННОЕ ИЗДАТЕЛЬСТВО
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
МОСКВА 1963

513
Ш 86

Юлий Анатольевич Шрейдер

Что такое расстояние?

М., Физматгиз, 1963 г., 76 стр. с илл.

Редактор *А. П. Баева.*

Техн. редактор *Л. Ю. Плакше.* Корректор *О. А. Сигал.*

Сдано в набор 23 II 1963 г. Подписано к печати 19/IV 1963 г. Бумага 84×108$^{1}/_{32}$. Физ. печ. л. 2,34. Услови. печ. л. 3,89. Уч.-изд. л. 3,72. Тираж 67 000 экз. Т-04942. Цена книги 11 коп. Заказ № 178.

Государственное издательство физико-математической литературы.
Москва, В-71, Ленинский проспект, 15.

Первая Образцовая типография имени А. А. Жданова
Московского городского совнархоза. Москва, Ж-54, Валовая, 28.

Отпечатано с матриц в гос. типографии «Вайздас».
Вильнюс, ул. Страздялио 1. Заказ № 1796

СОДЕРЖАНИЕ

Предисловие 4
§ 1. Как определяются математические понятия 7
§ 2. Расстояние в элементарной геометрии и его свойства . . 10
§ 3. Определение метрического пространства и расстояния . . 16
§ 4. Некоторые примеры метрических пространств 23
§ 5. Пространство сообщений 33
§ 6. Автоматическое исправление ошибок в сообщениях . . 44
§ 7. Расстояния и нормы в многомерном пространстве . . . 52
§ 8. Сглаживание ошибок экспериментальных измерений . . 65
§ 9. Более общие определения расстояния 69

ПРЕДИСЛОВИЕ

Эта книга является развитием лекции, прочитанной автором в Московском университете для школьников 9—10 классов. В ней рассказывается, как из простого геометрического понятия с помощью математической абстракции возникло общее определение расстояния. Приведены различные примеры пространств с расстоянием, так называемых метрических пространств. При этом оказывается, что общее понятие расстояния связано с разнообразными математическими фактами.

На основе понятия расстояния можно изучать задачи о кратчайших путях на поверхностях, геометрические свойства многомерных пространств, методы помехоустойчивого кодирования сообщений, методы «сглаживания» результатов измерений и др.

На примере «расстояния» видно, какую роль в математике играет создание общих понятий, находящих порой самые неожиданные применения и связи. Кроме «расстояния», можно было бы еще указать на понятия «функции», «предела», «пространства», «преобразования» и менее известные в широкой аудитории понятия «изоморфизма», «группы», «кольца» и др. Среди этих понятий «расстояние» — одно из наиболее доступных для элементарного объяснения, чем, пожалуй, в основном обусловлен выбор темы этой книги.

Автору хотелось доступными для массового читателя средствами показать, как одна плодотворная идея освещает широкий круг вопросов и служит источником для получения неожиданных результатов или нового взгляда на какую-либо область знания. Эта ситуация, характерная для любой науки, в математике очень часто проявляется в наиболее чистом виде, не заслоненная обилием необходимых, но мешающих подробностей.

Материал, отобранный для книги, в основном диктуется этим общим замыслом.

Первые четыре параграфа как раз и должны раскрыть читателю, как происходит переход от обычного геометрического определения к общему понятию «расстояния» и что это новое «понятие» означает в различных конкретных случаях.

В пятом параграфе описывается так называемое пространство сообщений, играющее важную роль в теории информации и общей теории связи.

Следующий параграф посвящен описанию методов кодирования сообщений, при которых сообщение оказывается устойчивым к ошибкам, возникающим в процессах передачи. Так как во всех реальных средствах связи время от времени возможны ошибки, то эти методы кодирования чрезвычайно существенны для современных систем связи и управления. Так, при передаче с борта космической ракеты на Землю фотографии обратной стороны Луны использовались помехоустойчивые методы кодирования этого сообщения. Необходимо отметить, что основная идея этих методов состоит в надлежащем использовании свойств расстояния в пространстве сообщений.

Седьмой параграф несколько более сложен для понимания. В нем описывается важный класс пространств с расстоянием.

В восьмом параграфе показано, как с помощью понятия расстояния можно получить методы сглаживания результатов измерений, т. е. фактически снижения ошибки в измерении каких-то опытных данных за счет соответствующей математической обработки. По существу в этом параграфе изложен так называемый метод наименьших квадратов. Для понимания этого параграфа нужны некоторые сведения из дифференциального исчисления. Читатель, не обладающий необходимыми познаниями, может этот параграф опустить.

В последнем параграфе рассматриваются возможные обобщения понятия расстояния. Здесь автору хотелось показать, что далеко не всякое обобщение является содержательным, т. е. обладающим интересными свойствами.

Придумать хорошее обобщение какого-либо математического понятия на самом деле не просто. В основе содержательного обобщения всегда лежат какие-то существенные свойства реального мира. В частности, важность понятия расстояния состоит в том, что целый ряд свойств многих реальных объектов связан с их взаимным расположением, которое часто можно охарактеризовать надлежащим образом определенным

расстоянием. Так, например, хотя электроны в оболочке атомов нельзя представлять себе в виде материальных точек, все же в квантовой механике можно особым образом определить «расстояние» между различными состояниями электронов в атоме. Это «расстояние» по идее близко к одному из определений расстояния из § 7 (так называемое пространство l_2).

Автор будет считать свою задачу выполненной удовлетворительно, если эта книга действительно раскроет перед читателем идеи, о которых говорилось выше.

Автор пользуется случаем выразить признательность И. М. Яглому, давшему ряд ценных советов по улучшению рукописи.

§ 1. КАК ОПРЕДЕЛЯЮТСЯ МАТЕМАТИЧЕСКИЕ ПОНЯТИЯ

Заголовок этой книги может вызвать недоумение. Ведь, что такое расстояние, знает, казалось бы, каждый старшеклассник. Да и человек, основательно забывший школьную геометрию и не умеющий четко сформулировать определение расстояния, скорее всего будет утверждать, что ему хорошо понятно, что это такое.

Вместе с тем дело обстоит гораздо сложнее.

Слово «расстояние» может иметь разный смысл в зависимости от того, о каком пространстве идет речь. Сейчас мы увидим, что так обстоит дело даже в привычных для нас ситуациях.

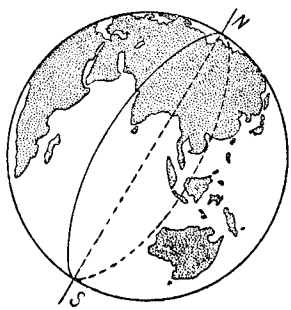

Рис. 1.

На плоскости и в обычном трехмерном (так называемом евклидовом) пространстве расстояние между двумя точками M и N определяется как длина отрезка MN, соединяющего эти точки. Однако когда мы имеем дело с расстояниями между географическими объектами, на поверхности земли, то мы обычно имеем в виду длину дуги большого круга, соединяющего эти объекты. Разница между этими двумя видами расстояний особенно наглядна, если вычислить расстояния между северным N и южным S полюсами (см. рис. 1). Обычное (евклидово) расстояние между полюсами равно диаметру земли, т. е. около 12 800 *км*. Расстояние между полюсами по поверхности земли больше первого в $\frac{\pi}{2}$ раз, т. е. равно примерно 20 100 *км*.

К этому примеру можно добавить, что в житейском обиходе при оценке расстояния между городами учитывается еще и способ передвижения. Так, расстояние по железной и шоссейной дороге между двумя пунктами может быть различным.

Новый пример расстояния мы получим, если будем рассматривать пункты на пересеченной местности, определяя расстояние между двумя пунктами как время, необходимое пешеходу для перехода из одного пункта в другой.

Ясно, что это расстояние не имеет ничего общего с длиной отрезка, соединяющего соответствующие пункты, так как путь по прямой, вообще говоря, не является наилучшим.

Вместе с тем любой путник исчисляет расстояние между пунктами своего маршрута по времени перехода.

Однако, несмотря на различие всех этих примеров, видно, что слово «расстояние» всюду употреблено в сходном смысле. Это слово везде означает меру удаленности каких-то объектов.

Итак, можно предположить, что есть некое общее определение расстояния, которое в разных конкретных ситуациях может иметь разное содержание. Такое общее определение будет сформулировано в § 3.

А сейчас мы обсудим, какие требования надо предъявлять к определению математического понятия.

Современная математика — это язык естествознания. В основе наиболее важных математических образов лежат какие-то пространственно-временные формы мира, в котором мы живем. Однако отношение между этими формами и математическими образами является очень сложным.

В любом разделе математики есть какие-то основные первоначальные понятия, которые в нашем сознании связаны с теми или иными физическими образами. Некоторые основные свойства этих понятий формулируются в виде аксиом (и постулатов), т. е. истин, которые не доказываются, а принимаются в качестве исходных. Все остальные положения данного раздела математики логически выводятся из этих аксиом без дополнительных ссылок на свойства физического мира. Сами формулировки аксиом отражают какие-то интуитивно очевидные свойства этих понятий, соответствующие очевидным свойствам их физических прообразов.

В геометрии такими основными понятиями являются: точка, прямая, плоскость, пространство и т. п. В систематических курсах геометрии можно найти перечень основных

свойств этих объектов, выраженных в виде так называемых аксиом и постулатов, на основе которых строится все здание геометрии *).

В алгебре такими основными понятиями являются множество чисел и действия над числами. В школе последовательно изучаются действия над целыми, дробными (или рациональными) относительными, вещественными (рациональными и иррациональными) числами и, наконец, комплексными числами.

При этом в каждом из перечисленных пяти случаев проверяется выполнимость основных законов (свойств действий над числами). Это — переместительный (коммутативный) закон для сложения ($a+b=b+a$), сочетательный (ассоциативный) закон для сложения ($(a+b)+c=a+(b+c)$), переместительный (коммутативный) закон для умножения ($ab=ba$), сочетательный (ассоциативный) закон для умножения ($(ab)c=a(bc)$), распределительный (дистрибутивный) закон $(a+b)c=ac+bc$ и правила: $a-a=0$; $a\cdot\frac{1}{a}=1$, характеризующие связь между основными (сложение и умножение) и обратными (вычитание, деление) операциями. Все эти законы выполняются во всех перечисленных числовых системах. Однако в некоторых из них деление выполнимо всегда, а в некоторых нет. Если числовая система состоит только из положительных чисел, то в ней не всегда возможно и вычитание. Оказывается, что правила алгебраических преобразований различных выражений основаны только на перечисленных выше свойствах. Все правила решения уравнений первой степени и систем таких уравнений также основаны на этих законах и возможности выполнения деления.

Оказывается, что свойства различных числовых систем можно изучать, исходя из общего понятия системы величин, над которыми определены операции (называемые сложением и умножением), удовлетворяющие перечисленным выше свойствам. Такие системы называются в современной алгебре *кольцами* или *полями* (в зависимости от того, можно ли всегда в этой системе выполнять деление) **).

Можно рассматривать правила преобразования выражений и правила решения уравнений в произвольном поле или кольце и как частный случай получить отсюда всю школьную алгебру.

---

*) Впервые так построил курс геометрии древнегреческий математик Евклид (IV—III век до н. э.).

**) Определение кольца и поля см. в книге: А. Г. Курош, Высшая алгебра, Москва, 1961, Физматгиз.

В современной алгебре рассматриваются кольца и поля, являющиеся обобщением изучаемых в школе числовых систем. При этих обобщениях в основу кладутся основные свойства действий, проверенные для целых или дробных чисел, и изучаются факты, вытекающие логически только из этих свойств.

При таком подходе математика интересует не только открытие новых свойств физического мира и установление связей между его свойствами, но и выяснение свойств «воображаемых» миров, вытекающих из каких-то положенных в основу представлений.

Эта черта математики не менее важна, чем возможность описания физического мира. Замечательный русский ученый Н. И. Лобачевский, изменив один из постулатов Евклида, создал так называемую «воображаемую» геометрию. Эта геометрия, спустя очень долгое время, легла в основу новых физических представлений о мире, возникших в результате открытия А. Эйнштейном теории относительности.

В этой книге мы рассматриваем одно из важных математических понятий — понятие «расстояния». Сначала мы проанализируем, какие основные законы, связанные с расстоянием, имеют место в элементарной геометрии. Положив в основу эти законы, мы получим определение так называемого «метрического пространства» и изучим различные примеры таких пространств. Мы увидим, что такой специфичный для математики подход к изучению различных объектов с точки зрения какого-то общего понятия открывает много интересных фактов.

Этот подход — создание новых общих понятий и попытка описать разнообразные объекты с помощью этих понятий — наиболее характерен для современной математики и ее приложений *). С этой точки зрения «расстояние» является удобным примером для того, чтобы убедиться в плодотворности такого подхода.

§ 2. РАССТОЯНИЕ В ЭЛЕМЕНТАРНОЙ ГЕОМЕТРИИ И ЕГО СВОЙСТВА

К общему определению расстояния мы придем, обобщая свойства обычного расстояния в трехмерном пространстве. Поэтому мы сначала сформулируем основные свойства обычного расстояния.

*) Нельзя не указать на роль, которую играют в кибернетике такие общие математические понятия, как информация, автомат, алгоритм.

Расстояние между двумя точками M и N в трехмерном пространстве — длину отрезка MN — принято обозначать $\varrho(M, N)$.

Это обозначение подчеркивает, что расстояние есть вещественное число, определяемое точками M и N. Иначе говоря, расстояние есть числовая функция от пары точек. Если каждую из точек характеризовать тройкой координат

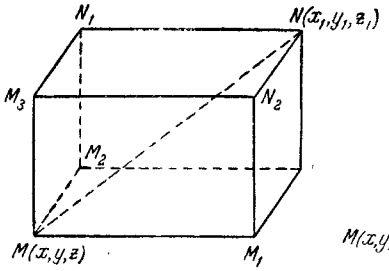

Рис. 2.

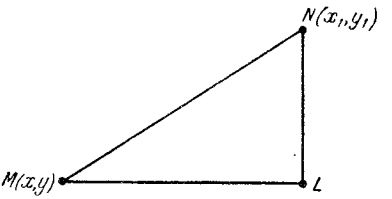

Рис. 3.

$M(x, y, z)$ и $N(x, y, z)$, то расстояние в пространстве есть функция шести переменных

$$\varrho(M, N) = F(x, y, z, x_1, y_1, z_1).$$

С помощью рис. 2 можно получить явное алгебраическое выражение для этой функции. На этом рисунке изображен параллелепипед со сторонами, параллельными осям координат. Известно, что квадрат диагонали параллелепипеда равен сумме квадратов диагоналей его сторон.

Следовательно,

$$MN^2 = MM_1^2 + MM_2^2 + MM_3^2 = (x-x_1)^2 + (y-y_1)^2 + (z-z_1)^2.$$

Отсюда

$$\varrho(M, N) = \sqrt{(x-x_1)^2 + (y-y_1)^2 + (z-z_1)^2}. \qquad (1)$$

Еще проще вычисляется расстояние между точками $M(x, y)$ и $N(x_1, y_1)$ на плоскости (см. рис. 3). Для этого надо заметить, что длина отрезка ML равна $|x-x_1|$, а длина отрезка LN равна $|y-y_1|$. По теореме Пифагора

$$MN^2 = ML^2 + LN^2,$$

откуда

$$\varrho(M, N) = \sqrt{(x-x_1)^2 + (y-y_1)^2}. \qquad (2)$$

Несмотря на важность соотношений (1) и (2), нужные нам свойства расстояния могут быть получены без использования координатной системы.

Эти свойства можно записать следующим образом:

1. $\varrho(M, N) = \varrho(N, M)$ (симметричность).
2. $\varrho(M, N) \geqslant 0$ (неотрицательность).
3. $\varrho(M, N) = 0$ в том и только в том случае, когда точки M и N совпадают (невырожденность).
4. $\varrho(M, N) \leqslant \varrho(M, L) + \varrho(L, N)$ — для любой тройки точек M, L, N (аксиома треугольника).

Свойства 1, 2 и 3 очевидны. Они означают просто, что длина отрезка MN не зависит от порядка, в котором берутся точки M и N, всегда неотрицательна и равна нулю, только если начало и конец отрезка MN совпадают.

Свойство 4 становится очевидным, если через точки M, N и L провести плоскость и рассмотреть в этой плоскости треугольник MLN (см. рис. 4). Тогда свойство 4 означает просто, что длина стороны MN не превосходит суммы длин остальных сторон треугольника MLN, откуда и происходит название свойства 4. Это свойство означает, что прямолинейный отрезок MN является кратчайшим путем, соединяющим точки M и N.

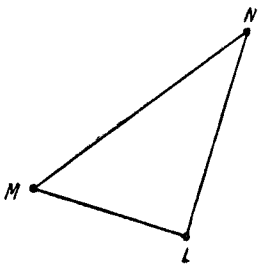

Рис. 4.

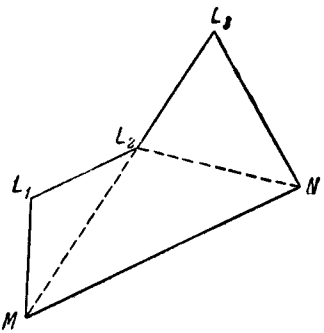

Рис. 5.

В самом деле, неравенство треугольника буквально означает, что длина отрезка MN меньше длины двухзвенной ломаной MLN, если точка L не лежит внутри этого отрезка. Отсюда можно вывести, что длина отрезка MN меньше длины

ломаной с любым числом звеньев, соединяющей точки M и N. Для этого (см. рис. 5) будем последовательно уменьшать число звеньев ломаной на единицу, спрямляя по два звена. Каждый раз при этом длина ломаной будет уменьшаться, пока мы не дойдем до отрезка MN. Так, на рис. 5 мы переходим от ломаной $ML_1L_2L_3N$ к ломаной ML_1L_2N, затем к ломаной ML_2N, а затем к отрезку MN. Каждый раз длина ломаной уменьшается, а, значит, длина исходной ломаной больше длины отрезка MN.

Отметим, что в этом рассуждении мы пользовались только неравенством треугольника. Действительно, составим такую таблицу:

Ломаная	Ее длина	Используемое неравенство треугольника
$ML_1L_2L_3N$	$\varrho(M, L_1)+\varrho(L_1, L_2)+\varrho(L_2, L_3)+\varrho(L_3, N)$	$\varrho(L_2, L_3)+\varrho(L_3, N)>\varrho(L_2, N)$
ML_1L_2N	$\varrho(M, L_1)+\varrho(L_1, L_2)+\varrho(L_2, N)$	$\varrho(M, L_1)+\varrho(L_1, L_2)>\varrho(M, L_2)$
ML_2N	$\varrho(M, L_2)+\varrho(L_2, N)$	$\varrho(M, L_2)+\varrho(L_2, N)>\varrho(M, N)$
MN	$\varrho(M, N)$	

Из этой таблицы видно, как, подставляя в суммы, стоящие во втором столбце, неравенства из третьего столбца, мы приходим, наконец, к выводу, что

$$\varrho(M, L_1)+\varrho(L_1, L_2)+\varrho(L_2, L_3)+\varrho(L_3, N)>\varrho(M, N).$$

Если воспользоваться дополнительно тем фактом, что длина кривой есть предел длин вписанных в эту кривую ломаных, то можно доказать следующее утверждение.

Среди всех линий, соединяющих точки M и N, отрезок MN имеет наименьшую длину.

Из неравенства треугольника следует, что

$$\varrho(L, N)\geqslant\varrho(M, N)-\varrho(M, L). \qquad (3)$$

Неравенство треугольника показывает, что если точка M близка к точке L, а точка L близка к точке N, то точки M и N тоже близки. Подчеркнем, что в свойстве 4 неравенство превращается в равенство в том и только том случае, когда

точки M, L и N лежат на одной прямой, а точка L находится между M и N (принадлежит отрезку MN).

Рассмотрим теперь расстояние на поверхности сферы S радиуса R.

Расстояние между двумя точками M и N на поверхности сферы определим как длину меньшей дуги большого круга, проходящей через точки M и N. Напомним, что большим кругом на сфере называется окружность, лежащая на поверхности сферы с центром в центре сферы.

Большой круг лежит в плоскости, проходящей через точки M, N и O (центр сферы). Отсюда следует, что для несовпадающих точек M и N большой круг определяется единственным образом, поскольку для него заданы центр и две точки на окружности. Определенное таким образом расстояние $\varrho_S(M, N)$, очевидно, удовлетворяет свойствам 1, 2 и 3. Нетрудно также видеть, что для любых точек на сфере

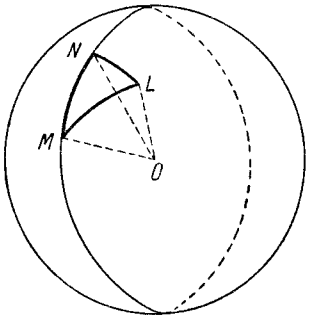

Рис. 6.

$$\varrho_S(M, N) \leqslant \pi R, \qquad (4)$$

причем равенство достигается только для точек M и N, лежащих на концах общего диаметра сферы (например, северный и южный полюсы).

Для того чтобы проверить свойство 4, надо рассмотреть сферический треугольник MLN (см. рис. 6). На этом рисунке точка O есть центр сферы.

Ясно, что

$$\varrho_S(M, N) = R\alpha,\ \varrho_S(L, N) = R\beta,\ \varrho_S(L, M) = R\gamma,$$

где

$$\alpha = \angle MON;\ \beta = \angle LON,\ \gamma = \angle LOM.$$

Известно, что в трехгранном угле каждый плоский угол не превосходит суммы двух других плоских углов, т. е.

$$\alpha \leqslant \beta + \gamma. \qquad (5)$$

Умножая обе части этого неравенства на радиус R, получим

$$R\alpha \leqslant R\beta + R\gamma.$$

Это и означает, что

$$\varrho_S(M, N) \leqslant \varrho_S(M, L) + \varrho_S(L, N), \qquad (6)$$

что и требовалось доказать.

Таким образом, для сферического расстояния $\varrho_S(M, N)$ выполнены все основные свойства обычного расстояния.

Нетрудно проверить, что неравенство (6) обращается в равенство, только когда точка L расположена на том же большом круге, что и точки M и N. При этом точка L должна находиться между точками M и N, т. е. на меньшей дуге большого круга, соединяющей точки M и N.

Это следует из того, что неравенство (6) обращается в равенство только тогда, когда в равенство обращается (5). Последнее же может произойти, только когда трехгранный угол вырождается в плоский, т. е. когда точки M, L и N лежат в одной плоскости с центром сферы O и луч OL находится между лучами OM и ON. Но это и значит, что точка L принадлежит меньшей дуге большого круга, соединяющего точки M и N.

Отсюда видно, что меньшая дуга большого круга, соединяющая точки M и N, обладает свойствами, аналогичными свойствам отрезка в обычной (не сферической) геометрии. А именно: 1) через любые не совпадающие точки проходит ровно одна такая дуга (исключением является случай, когда точки M и N лежат на концах некоторого диаметра сферы — являются антиподами; в этом случае обе дуги любого большого круга, соединяющего M и N, равноправны, а таких кругов бесконечно много); 2) две такие дуги могут пересекаться только в одной точке; 3) для любой точки L, лежащей на такой дуге, соединяющей точки M и N, имеет место равенство $\varrho_S(M, L) + \varrho_S(L, N) = \varrho_S(M, N)$.

Заметим теперь следующее важное обстоятельство. Для обычного расстояния мы показали, что длина ломаной, соединяющей две точки M и N, больше, чем расстояние между точками M и N, т. е. чем длина отрезка MN. При этом рассуждение основывалось только на неравенстве треугольника и том факте, что оно обращается в равенство, лишь когда точки M, L и N лежат на одном отрезке. Так как неравенство треугольника верно и для расстояния на сфере, а обычным отрезкам здесь соответствуют меньшие дуги больших кругов, то ясно, что аналогичное утверждение верно и на сфере. Именно, если точки M и N соединены цепочкой из дуг больших кругов (см. рис. 7), последовательно

соединенных друг с другом, то суммарная длина такой «сферической ломаной» больше *), чем расстояние $\varrho_S(M, N)$.

Можно предложить читателю для упражнения восстановить полное доказательство этого утверждения по аналогии с приведенным выше доказательством для обычной ломаной. Это утверждение можно легко обобщить в следующей форме. *Длина меньшей дуги большого круга, соединяющей точки M и N, меньше, чем длина любой другой линии на сфере, соединяющей эти точки.*

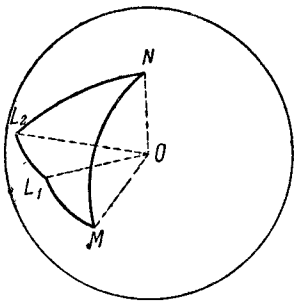

Рис. 7.

Итак, мы рассмотрели два вида «расстояний» и убедились, что их основные свойства совпадают. Дополнительные свойства типа (4) играют гораздо меньшую роль.

Поэтому следующий наш шаг будет состоять в том, что мы основные свойства расстояния (1, 2, 3, 4) возьмем в качестве исходных и будем изучать различные «пространства», в которых определено «расстояние», обладающее этими исходными свойствами. В этом параграфе мы рассмотрели два простейших примера таких пространств: обычное пространство и поверхность сферы.

§ 3. ОПРЕДЕЛЕНИЕ МЕТРИЧЕСКОГО ПРОСТРАНСТВА И РАССТОЯНИЯ

Мы начнем с разъяснения того, что такое множество. Так же как «точка» в геометрии, понятие «множества» является первоначальным и определению не подлежит. Под словом «множество» в математике понимается совокупность каких-то объектов, называемых элементами множества.

Понятие «множества» состоит в сущности в том, что каким-то объектам приписывается свойство общности. Иначе, эти объекты по каким-то признакам попадают в некий класс, т. е. образуют множество. Задать множество — это и значит задать классификацию, т. е. для любого мыслимого объекта иметь возможность определить, попадает он в этот класс

*) Разумеется, если эта цепочка не лежит целиком на меньшей дуге большого круга, соединяющей точки M и N.

или нет. Принято говорить, что множество *содержит* каждый из своих элементов, а каждый из элементов данного множества в нем содержится. Множество считается заданным, если для любого объекта можно как-то проверить, содержится ли он в данном множестве *).

Можно рассматривать, например, множество всех целых чисел. Солнце не содержится в этом множестве, так как оно не есть число, а объект совсем другой природы. Число π также не содержится в этом множестве, ибо оно не целое. С другой стороны, корни уравнения $x^2-3x+2=0$ входят в это множество. Можно рассматривать множество всех планет солнечной системы, понимая под планетами тела, движущиеся вокруг Солнца по замкнутым орбитам и имеющие массу не меньше чем 1 тонна. Солнце не содержится в этом множестве, так как является звездой, а не планетой. Земля содержится в этом множестве. Советская ракета, запущенная с Земли 2 января 1959 года, содержится в этом множестве, так как является искусственной планетой.

Пусть E — некоторое множество, а N — его элемент. Этот факт символически записывается как $N \in E$. Читается: «N является элементом E». Применяется также символическая запись типа $E\{L, M, N, \ldots\}$, где в скобках перечислены все элементы этого множества. Так, множество E_c, состоящее из всех столиц союзных республик, можно было бы записать символически так: E_c {Москва, Киев, Минск, Тбилиси, Ереван, Баку, Рига, Таллин, Вильнюс, Ташкент, Алма-Ата, Фрунзе, Ашхабад, Дюшамбе, Кишинев}.

Если любой элемент множества E является в то же время элементом множества E_1, то множество E называется *подмножеством* множества E_1. Это записывается так: $E \subset E_1$ («E содержится в E_1»).

Так, множество целых чисел является подмножеством множества вещественных чисел.

Множество E называется *конечным*, если его элементы можно перенумеровать с помощью некоторого отрезка натурального ряда — множества $E_n\{1, 2, 3, \ldots, n\}$. Например,

*) Вопрос о том, какой способ проверки считать эффективным, предст вляет большой интерес для математической логики или философии, но здесь нас интересовать не будет. Имеющиеся здесь трудности аналогичны трудностям любой формальной классификационной системы. Примером тому могут служить описанные в книге Веркора «Люди или животные» трудности в определении, какие человекообразные принадлежат виду Homo sapiens.

множество столиц союзных республик E_c конечно, так как его можно перенумеровать элементами множества E_{15}, что видно из следующей таблицы соответствий:

Москва 1	Ереван 5	Вильнюс 9	Ашхабад 13
Киев 2	Баку 6	Ташкент 10	Дюшамбе 14
Минск 3	Рига 7	Алма-Ата 11	Кишинев 15
Тбилиси 4	Таллин 8	Фрунзе 12	

Теперь мы можем дать определение метрического пространства.

Метрическим пространством называется такое множество E, что для любой пары его элементов M и N определено вещественное число $\varrho(M, N)$, обладающее следующими свойствами:

1. $\varrho(M, N) = \varrho(N, M)$ (симметричность).
2. $\varrho(M, N) \geqslant 0$ (неотрицательность).
3. $\varrho(M, N) = 0$ в том и только в том случае, когда M и N один и тот же элемент (невырожденность).
4. $\varrho(M, N) \leqslant \varrho(M, L) + \varrho(L, N)$ для любой тройки M, L, N элементов множества E (аксиома треугольника).

Элементы множества E принято называть *точками* соответствующего *пространства*. Метрическое пространство определяется, таким образом, выбором множества E и функции $\varrho(M, N)$ — *расстояния в пространстве*. Для простоты будем обозначать получаемое пространство той же буквой, что и соответствующее множество, хотя на самом деле пространство и множество его элементов суть разные объекты. Действительно, в одном и том же множестве E можно ввести различные расстояния и получить разные метрические пространства. В § 4 мы построим различные определения расстояния на плоскости.

Вместо указанных четырех свойств расстояния можно было бы ввести только два (предполагая заранее, что $\varrho(M, N)$ — вещественное число):

1′. $\varrho(M, N) = 0$ тогда и только тогда, когда точки M и N совпадают.

2′. $\varrho(M, N) \leqslant \varrho(M, L) + \varrho(N, L)$.

Действительно, эти свойства вытекают из свойств 1, 2, 3 и 4, так как свойство 1′ совпадает со свойством 3, а свойство 2′ вытекает из неравенства треугольника 4 и условия 1.

С другой стороны, из свойств 1′ и 2′ можно вывести все условия 1, 2, 3 и 4.

Покажем это. Положим сначала в свойстве 2′ $L=M$, тогда получим:

$$\varrho(M,\ N)\leqslant\varrho(M,\ M)+\varrho(N,\ M).$$

В силу 1′ $\varrho(M,\ M)=0$. Отсюда следует $\varrho(M,\ N)\leqslant\varrho(N,\ M)$. Меняя в 2′ местами точки M и N и проводя аналогичное рассуждение, убеждаемся, что $\varrho(N,\ M)\leqslant\varrho(M,\ N)$. Из последних двух неравенств следует свойство симметрии (1):

$$\varrho(M,\ N)=\varrho(N,\ M).$$

Подставляя в 2′ M вместо N и L, получаем

$$\varrho(M,\ M)\leqslant\varrho(M,\ M)+\varrho(N,\ M)$$

или

$$0\leqslant\varrho(N,\ M),$$

что доказывает условие неотрицательности расстояния (2). Наконец, используя доказанное условие симметрии, можно в 2′ поменять во втором слагаемом правой части N и L и получить неравенство треугольника 4. Таким образом, система свойств 1′ и 2′ равносильна системе свойств 1, 2, 3 и 4. Удобнее пользоваться более полной системой, которая в явном виде содержит больше фактических свойств расстояния. Однако интересно знать, что все эти свойства могут быть выведены всего из двух.

С точки зрения введенного определения содержание предыдущего параграфа можно было бы охарактеризовать как доказательство, что множество точек трехмерного пространства с расстоянием, определенным как длина отрезка или по формуле (1), является метрическим пространством. Аналогичный факт доказан в конце того же параграфа про множество точек сферы S с расстоянием $\varrho_S(M,\ N)$.

Еще один пример метрического пространства получится, если взять множество точек некоторой поверхности Π в трехмерном пространстве, а в качестве расстояния $\varrho_\Pi(M,\ N)$ взять минимум длин путей, проходящих по поверхности Π и соединяющих точки M и N*). Первые три свойства расстояния очевидны.

Аксиома треугольника проверяется так. Соединим точки M и L наикратчайшим путем, и точки L и N также соединим

*) Для простоты мы предполагаем, что такой наикратчайший путь существует для любых точек M и N на поверхности Π. При некоторых предположениях о свойствах поверхности Π это можно доказать.

наикратчайшим путем. Тогда путь, состоящий из двух наикратчайших линий ML и LN, соединяет точки M и N. Ясно, что длина этого пути не может быть меньше длины наикратчайшего пути из M в N. Так как длина этого пути есть $\varrho_{\text{П}}(M, L)+\varrho_{\text{П}}(L, N)$, а длина наикратчайшего пути из M в N есть $\varrho_{\text{П}}(M, N)$, то отсюда следует искомое соотношение:

$$\varrho_{\text{П}}(M, N)\leqslant\varrho_{\text{П}}(M, L)+\varrho_{\text{П}}(L, N). \tag{7}$$

Заметим, что на поверхности сферы наикратчайший путь между точками есть как раз наименьшая дуга большого круга, соединяющая эти точки. Доказательство этого факта было намечено в конце предыдущего параграфа. Это доказательство как раз и основано на том, что независимыми рассуждениями было получено неравенство треугольника и показано, что для наименьших дуг большого круга оно обращается в равенство.

Полезно ввести понятие отрезка в произвольном метрическом пространстве. Именно, мы назовем *отрезком, соединяющим точки M и N в метрическом пространстве E*, такое множество точек $E_{M,N}$ этого пространства, что для любой точки L множества $E_{M,N}$ выполнено равенство

$$\varrho(M, N)=\varrho(M, L)+\varrho(L, N). \tag{8}$$

Нетрудно видеть, что для обычного расстояния на плоскости или в трехмерном пространстве множество $E_{M,N}$ совпадает с отрезком MN в обычном смысле слова. На сфере S при расстоянии $\varrho_S(N, M)$, введенном в § 2, отрезком $E_{M,N}$ является наименьшая дуга большого круга, соединяющая точки M и N (если они не лежат на одном диаметре) и весь большой круг, если точки M и N — антиподы на сфере.

Предоставляем читателю убедиться, что для расстояния $\varrho_{\text{П}}(M, N)$, введенного выше, отрезок $E_{M,N}$ есть кратчайший путь (или так называемая геодезическая линия), соединяющий точки M и N.

В любом метрическом пространстве E можно также ввести понятие сферы $S_{M,r}$ с центром в точке M и радиусом r, обозначая этим термином множество таких точек N, для которых расстояние $\varrho(M, N)=r$.

На плоскости это понятие соответствует окружности; в трехмерном пространстве — обычной сфере; для расстояния $\varrho_S(M, N)$ — окружностям на сфере S.

В качестве еще одного (тривиального) примера метрического пространства возьмем произвольное множество E и определим расстояние между двумя любыми не совпадающими точками M и N равным единице, а между совпадающими точками — нулю. Легко видеть, что все необходимые условия при таком определении выполнены.

Разнообразные примеры метрических пространств будут рассмотрены в § 4.

В метрическом пространстве можно всегда определить операцию предельного перехода.

Последовательность точек метрического пространства E: $L_1, L_2, \ldots, L_k$ называется *сходящейся* к точке $L \in E$, если для всякого положительного вещественного числа ε можно подобрать такой номер $n(\varepsilon)$, что из условия $k > n(\varepsilon)$ следует

$$\rho(L, L_k) < \varepsilon.$$

Иными словами, последовательность сходится к точке L, если, начиная с некоторого номера, расстояние между членами последовательности и точкой L (пределом) становится меньше любого наперед заданного положительного числа. По аналогии с обычной записью можно ввести обозначение

$$L = \lim_{k \to \infty} L_k.$$

Нетрудно убедиться, что для метрического пространства, состоящего из вещественных чисел x, с расстоянием, определяемым по правилу

$$\rho(x, y) = |x - y|,$$

это определение предела совпадает с обычным.

Для метрического пространства E_3, состоящего из точек трехмерного пространства с обычным расстоянием, данное понятие предела позволяет хорошо определить, что такое предел последовательности точек в трехмерном пространстве.

Заметим, что в этом случае совокупность точек M, для которых $\rho(L, M) < \varepsilon$, образует сферу с центром в точке L и радиусом ε. Последовательность геометрических точек $L_1, L_2, \ldots, L_k, \ldots$ оказывается *сходящейся* к точке L, если, начиная с некоторого номера $n(\varepsilon)$, все члены этой последовательности с $k > n(\varepsilon)$ лежат в сфере с центром в L и радиусом ε, причем такой номер $n(\varepsilon)$ должен существовать для всякого положительного ε.

Теорема. *Если последовательность элементов метрического пространства E: $L_1, L_2, \ldots, L_k, \ldots$ сходится к пределу L, то для всякого положительного числа ε существует такой номер $m(\varepsilon)$, что при $k > m(\varepsilon)$ и $k' > m(\varepsilon)$ выполняется условие $\rho(L_k, L_{k'}) < \varepsilon$.*

Доказательство. На основании определения предела можно выбрать такой номер $n\left(\frac{\varepsilon}{2}\right)$, что при $k > n\left(\frac{\varepsilon}{2}\right)$

и $k'>n\left(\frac{\varepsilon}{2}\right)$ будут справедливы неравенства

$$\varrho(L_k,\ L)<\frac{\varepsilon}{2};\quad \varrho(L_{k'},\ L)<\frac{\varepsilon}{2}.$$

Но по аксиоме треугольника и свойству симметричности

$$\varrho(L_k,\ L_{k'})\leqslant\varrho(L_k,\ L)+\varrho(L,\ L_{k'})=$$

$$=\varrho(L_k,\ L)+\varrho(L_{k'},\ L)<\frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon.$$

Иными словами, если положить $m(\varepsilon)=n\left(\frac{\varepsilon}{2}\right)$, то при $k>m(\varepsilon)$ и $k'>m(\varepsilon)$ будет иметь место неравенство

$$\varrho(L_k,\ L_{k'})<\varepsilon,$$

что и доказывает теорему.

Эта теорема позволяет охарактеризовать внутренние свойства сходящихся последовательностей.

Если в пространстве E справедливо утверждение, обратное этой теореме, то пространство E называется *полным* метрическим пространством.

Определение полного метрического пространства удобно дать в следующей развернутой форме. Последовательность точек метрического пространства E: $L_1, L_2, \ldots, L_k, \ldots$ называется *фундаментальной*, если для всякого числа $\varepsilon>0$ найдется номер $m(\varepsilon)$, такой, что при $k>m(\varepsilon)$ и $k'(\varepsilon)$ выполнено условие

$$\varrho(L_k,\ L_{k'})<\varepsilon.$$

Метрическое пространство E называется *полным*, если в нем всякая фундаментальная последовательность имеет предел.

Прямая, плоскость и трехмерное пространство с обычным расстоянием являются полными метрическими пространствами.

Вопрос о полноте метрических пространств очень существен при использовании этих пространств в математическом анализе, но здесь мы его больше касаться не будем *).

Два метрических пространства называются *изометрическими*, если между множествами их элементов можно установить взаимно однозначное соответствие, так что расстояния между парами сопоставленных друг другу элементов равны. С точки зрения теории метрических пространств изометрические пространства являются одинаковыми.

Рассмотрим пример. Пусть пространство E есть плоскость с обычным расстоянием, а пространство E_K есть множество

*) Для математического анализа принципиальное значение имеет полнота ряда метрических пространств, точками которых являются функции. См., например, определение пространства C в конце § 7.

комплексных чисел z с расстоянием, определяемым по формуле

$$\varrho_K(z, z_1) = |z - z_1|.$$

Обычный способ изображения комплексных чисел точками плоскости осуществляет взаимно однозначное соответствие между этими пространствами. Нетрудно проверить, что это соответствие является изометричным, так как величина $|z - z_1| = \sqrt{(x - x_1)^2 + (y - y_1)^2}$ равна геометрическому расстоянию между соответствующими точками.

Данное нами определение метрического пространства с расстоянием не является наиболее общим, встречающимся в математике. Существуют различные обобщения этого понятия. Так, можно было бы, сохраняя все свойства расстояния, разрешить ему принимать для некоторых точек бесконечные значения. Это обобщение, как мы убедимся в § 9, не очень содержательно. В ряде математических задач приходится иметь дело с расстоянием, у которого отсутствует свойство симметрии. Свойства такого расстояния мы изучим в § 9. Наконец, в теории относительности приходится рассматривать расстояние, которое может принимать и мнимые значения. Свойства такого расстояния довольно своеобразны, и мы их не будем затрагивать в этой книге.

§ 4. НЕКОТОРЫЕ ПРИМЕРЫ МЕТРИЧЕСКИХ ПРОСТРАНСТВ

В этом параграфе мы рассмотрим ряд примеров метрических пространств с необычно определяемым расстоянием.

Ряд интересных примеров получается, если пространство есть множество точек обычной плоскости, а расстояние вводится различными способами. Точки плоскости мы будем при этом изображать с помощью раз и навсегда выбранной координатной системы, так что каждая точка плоскости задается двумя координатами x и y. Поэтому точку плоскости удобно обозначать как $M(x, y)$.

Метрическое пространство l возникает, если расстояние между точками $M(x, y)$ и $N(x, y)$ определить по формуле

$$\varrho_1(M, N) = |x - x_1| + |y - y_1|. \qquad (9)$$

Из рис. 3 (стр. 11) можно усмотреть, что $\varrho_1(M, N)$ равно сумме длин катетов треугольника MLN, в котором MN — гипотенуза, а катеты ML и LN параллельны осям координат.

Так как длина гипотенузы не превосходит суммы длин ка тетов, то всегда

$$\varrho(M, N) \leqslant \varrho_1(M, N), \tag{10}$$

где $\varrho(M, N)$ есть обычное геометрическое расстояние. Нера венство (10) превращается в равенство в том и только в том случае, когда отрезок MN или вертикален, или гори зонтален, т. е. параллелен одной из осей координат.

Если в неравенство (10) подставить алгебраические выра жения для соответствующих расстояний (9) и (2), то мы получим неравенство

$$\sqrt{(x-x_1)^2+(y-y_1)^2} \leqslant |x-x_1|+|y-y_1|.$$

При $x_1=y_1=0$ отсюда следует простое, но красивое неравенство

$$\sqrt{x^2+y^2} \leqslant |x|+|y|. \tag{11}$$

Свойства 1, 2 и 3 для расстояния $\varrho_1(M, N)$ очевидны. Докажем, что выполнено и свойство 4. Для этого рассмотрим три точки $M(x, y)$, $N(x_1, y_1)$, $L(x_2, y_2)$ и запишем элемен тарное тождество

$$|x-x_1|+|y-y_1| = \\ =|x-x_2+x_2-x_1|+|y-y_2+y_2-y_1|. \tag{12}$$

Если использовать тот факт, что для любых вещественных чисел a и b, $|a+b| \leqslant |a|+|b|$, то из (12) получается не равенство

$$|x-x_1|+|y-y_1| \leqslant \\ \leqslant |x-x_2|+|x_2-x_1|+|y-y_2|+|y_2-y_1|,$$

что уже после перестановки членов в правой части равно сильно искомому соотношению

$$\varrho_1(M, N) \leqslant \varrho_1(M, L)+\varrho_1(L, N). \tag{13}$$

Тем самым доказано неравенство треугольника для про странства l.

Расстояние $\varrho_1(M, N)$ можно трактовать как минимальный путь, проходимый частицей, движущейся из M в N, если эта частица может двигаться только по отрезкам, параллельным осям координат.

Из рис. 8 видно, что существует много (бесконечно много!) наикратчайших путей для этой частицы.

Нетрудно показать, что это равносильно тому, что в пространстве l существует, вообще говоря, бесконечное множество различных «отрезков»*), соединяющих точки M и N (исключением является случай, когда точки M и N находятся на одной вертикали или горизонтали).

«*Отрезком*» в пространстве l, соединяющим точки M и N, является любая ломаная, соединяющая эти точки, состоящая

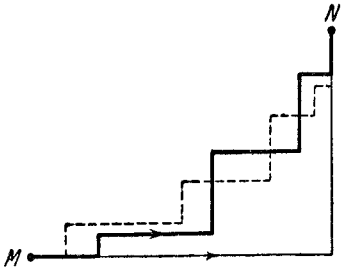

Рис. 8.

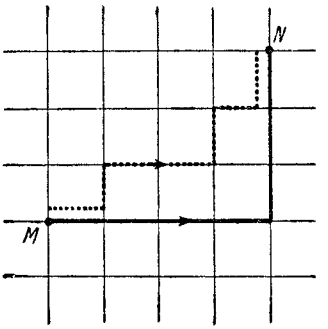

Рис. 9.

из вертикальных и горизонтальных линий и пересекаемая любой вертикальной или горизонтальной линией не более чем по одному звену. (Это утверждение предоставляем проверить читателю.)

Еще более естественная картина получится, если рассмотреть метрическое пространство l, состоящее из всех точек, находящихся в узлах некоторой прямоугольной решетки на плоскости (рис. 9) с расстоянием, определенным формулой (9). Точки этого пространства можно рассматривать как перекрестки улиц бесконечного идеально распланированного города (тоже своего рода математическая абстракция!). Расстояние $\varrho_1(M, N)$ есть в этом случае минимальная длина пути, который необходимо пройти по улицам города, чтобы попасть из перекрестка M в перекресток N (не пользуясь, разумеется, проходными дворами).

Следующий пример метрического пространства — пространство C состоит из точек плоскости с расстоянием,

*) В смысле введенного в § 3 определения.

определяемым по формуле

$$\varrho_\infty(M,\ N) = \max(|x-x_1|,\ |y-y_1|), \tag{14}$$

если точка M имеет координаты x и y, а точка N—координаты x_1 и y_1*). Геометрически это означает (см. рис. 3), что расстояние $\varrho_\infty(M,\ N)$ равно максимальному катету треугольника MLN. Так как даже максимальный катет меньше гипотенузы (или равен ей в вырожденном треугольнике), то

$$\varrho_\infty(M,\ N) \leqslant \varrho(M,\ N), \tag{15}$$

где $\varrho(M,\ N)$—обычное геометрическое расстояние.

Отсюда при $x_1 = y_1 = 0$ получается алгебраическое неравенство

$$\max(|x|,\ |y|) \leqslant \sqrt{x^2+y^2}. \tag{16}$$

Для расстояния $\varrho_\infty(M,\ N)$ также очевидны свойства 1, 2, 3. Аксиома треугольника проверяется следующим образом. Вы берем любые три точки $M(x,\ y)$, $N(x_1,\ y_1)$ и $L(x_2,\ y_2)$. Пред положим, что $|x-x_1| \geqslant |y-y_1|$**). Это значит, что

$$\varrho_\infty(M,\ N) = \max(|x-x_1|,\ |y-y_1|) = |x-x_1| = \\ = |x-x_2+x_2-x_1|.$$

Следовательно,

$$\varrho_\infty(M,\ N) \leqslant |x-x_2| + |x_2-x_1|. \tag{17}$$

Очевидно также, что

$$\left.\begin{aligned} |x-x_2| &\leqslant \max(|x-x_2|,\ |y-y_2|) = \varrho_c(M,\ L), \\ |x_2-x_1| &\leqslant \max(|x_2-x_1|,\ |y_2-y_1|) = \varrho_c(L,\ N). \end{aligned}\right\} \tag{18}$$

Подставляя правые части неравенств (18) в неравенство (17), получаем

$$\varrho_\infty(M,\ N) \leqslant \varrho_\infty(M,\ L) + \varrho_\infty(L,\ N), \tag{19}$$

что и требовалось доказать.

На примере города, изображенного на рис. 9, расстоянию $\varrho_\infty(M,\ N)$ можно придать следующий смысл. Предположим,

---

*) Смысл индекса ∞ будет раскрыт на стр. 29.

**) В противном случае доказательство видоизменяется очевидным образом.

что некто добирается до перекрестка N, отсчитывая по порядку номера продольных и поперечных улиц. Предположим, что длины всех кварталов одинаковы (и равны единице), и путешественник может сбиться со счета не более чем на k улиц (как продольных, так и поперечных).

Тогда путешественник может попасть на перекресток M, отстоящий от N на расстоянии $\varrho_\infty(M, N) \leqslant k$.

Мы уже указывали, что в любом метрическом пространстве можно ввести понятие обобщенной сферы. (Слово «обобщенный» мы в дальнейшем будем опускать.) Аналогично этому обобщенным шаром с центром в точке M и радиусом r мы будем называть множество точек N, для которых

$$\varrho(M, N) \leqslant r. \qquad (20)$$

Если $\varrho(M, N)$ — геометрическое расстояние на плоскости, то обобщенный шар является кругом с центром в точке M и радиусом r.

Для трехмерного пространства с обычным расстоянием шар, определяемый неравенством (20), есть обычный шар с центром в точке M и радиусом r. (Отсюда и происходит название «обобщенный шар».)

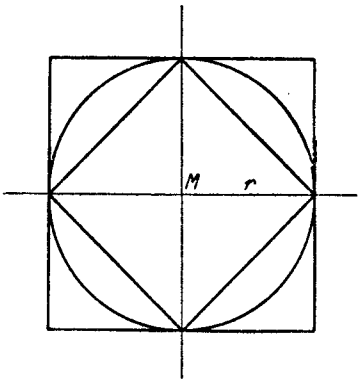

Рис. 10.

В пространстве l обобщенным шаром является квадрат с центром в точке M, с диагоналями, равными $2r$ и параллельными осям координат.

В пространстве C шаром является также квадрат с центром в точке M, но со сторонами, равными $2r$ и параллельными осям координат. На рис. 10 изображены обобщенные шары радиуса r в пространствах l и C и в смысле обычного расстояния. Доказательство того, что шары в l и C имеют указанный вид, предоставляется читателю.

Интересный класс метрических пространств получается, если определить расстояние $\varrho_p(M, N)$ по формуле

$$\varrho_p(M, N) = \sqrt[p]{|x-x_1|^p + |y-y_1|^p}, \qquad (21)$$

и получить тем самым пространство l_p.

Свойства 1, 2, 3 для этого расстояния очевидны. Аксиома треугольника следует из неравенства Минковского *)

$$\sqrt[p]{|a+a_1|^p+|b+b_1|^p} \leqslant \leqslant \sqrt[p]{|a|^p+|b|^p}+\sqrt[p]{|a_1|^p+|b_1|^p}, \quad (22)$$

справедливого при $p \geqslant 1$. Читатель легко получит из (22) доказательство аксиомы треугольника для расстояния (21) при $p \geqslant 1$, если для точек $M(x, y)$, $N(x_1, y_1)$ и $L(x_2, y_2)$ обозначит

$$a = x - x_2; \quad a_1 = x_2 - x_1; \quad b = y - y_2; \quad b_1 = y_2 - y_1$$

и будет действовать по аналогии с доказательством аксиомы треугольника для пространства l. При $p<1$ аксиома треугольника неверна, так как знак неравенства (22) меняется на противоположный.

Нетрудно видеть, что при $p=1$ расстояние $\varrho_p(M, N) = \varrho_l(M, N)$, а при $p=2$ расстояние $\varrho_p(M, N)$ совпадает с обычным геометрическим расстоянием $\varrho(M, N)$.

Таким образом, рассмотренное ранее пространство l совпадает с пространством l_1, а плоскость с обычным расстоянием есть пространство l_2.

Покажем, что при $p \longrightarrow \infty$ расстояние $\varrho_p(M, N)$ стремится к расстоянию $\varrho_\infty(M, N)$.

В самом деле, рассмотрим сначала случай $|x-x_1| > |y-y_1|$. Тогда $\varrho_\infty(M, N) = |x-x_1|$. С другой стороны, преобразуя (21), имеем

$$\varrho_p(M, N) = |x-x_1| \sqrt[p]{1+\left|\frac{y-y_1}{x-x_1}\right|^p}. \quad (23)$$

Замечая, что при $p>1$

$$1 \leqslant \sqrt[p]{1+\left|\frac{y-y_1}{x-x_1}\right|^p} \leqslant 1+\left|\frac{y-y_1}{x-x_1}\right|^p$$

и что величина $\left|\frac{y-y_1}{x-x_1}\right|^p \longrightarrow 0$ при $p \longrightarrow \infty$, получаем

$$\lim_{p\to\infty} \sqrt[p]{1+\left|\frac{y-y_1}{x-x_1}\right|^p} = 1.$$

*) Доказательство этого неравенства можно найти в книге: Г. Харди, Дж. Литлвуд и Полиа, Неравенства, перев. с англ., ИЛ, Москва, 1948.

Учитывая этот факт и используя (23), видим, что

$$\lim_{p\to\infty}\varrho_p(M, N)=|x-x_1|=\varrho_\infty(M, N). \tag{24}$$

Аналогично, при $|x-x_1|<|y-y_1|$ можно получить

$$\lim_{p\to\infty}\varrho_p(M, N)=|y-y_1|=\varrho_\infty(M, N). \tag{25}$$

Наконец, рассмотрим случай $|x-x_1|=|y-y_1|$. Тогда

$$\varrho_p(M, N)=\sqrt[p]{2|x-x_1|^p}=|x-x_1|\sqrt[p]{2}.$$

Так как $\lim\limits_{p\to\infty}\sqrt[p]{2}=1$, то и в этом случае

$$\lim_{p\to\infty}\varrho_p(M, N)=|x-x_1|=\varrho_\infty(M, N). \tag{26}$$

Сопоставляя (24), (25) и (26), получаем, что во всех случаях

$$\lim_{p\to\infty}\varrho_p(M, N)=\varrho_\infty(M, N), \tag{27}$$

что и требовалось доказать.

Таким образом, пространство C можно было бы обозначить символом l_∞, так как расстояние в этом пространстве получается предельным переходом из расстояния в пространстве l_p при $p\to\infty$.

На рис. 11 показаны шары в пространствах l_p с центром в точке M при различных значениях p.

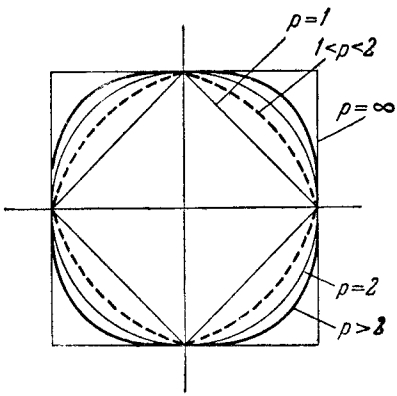

Рис. 11.

Метрические пространства l_p носят общее название пространств Минковского. В § 7 мы рассмотрим многомерные пространства Минковского.

В качестве упражнения для читателя предлагается найти вид «отрезков» в пространствах l_p.

Следующий класс метрических пространств на плоскости получится, если расстоянием называть минимальное время,

требующееся для прохождения при некоторых заданных условиях пути из M в N.

При этом обычное расстояние получается, если путь от M к N проходится точкой, движущейся с постоянной скоростью, равной единице.

Расстояние в l получится, если этой точке разрешается двигаться и с постоянной скоростью только по отрезкам, параллельным осям координат.

Новый пример получится, если рассмотреть (см. рис. 12) схему московского метрополитена и считать, что путешественник проходит из точки M в N следующим образом.

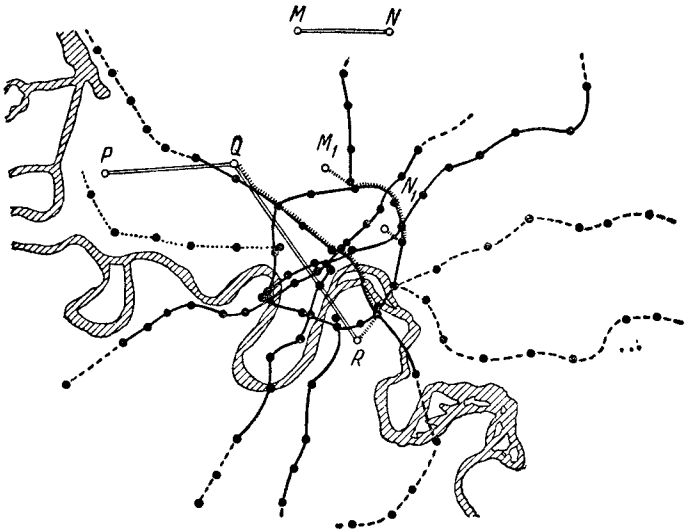

Рис. 12.

Если существует одна и та же ближайшая от обеих точек станция метро, то путешественник идет пешком по наикратчайшему пути. Если такой станции нет, то путешественник идет пешком (кратчайшим путем) до станции, ближайшей к исходной точке M, едет скорейшим путем к станции, ближайшей к точке N, и оттуда идет пешком к N. Если для точек M или N есть несколько ближайших станций метро, то выбирается вариант маршрута с наименьшим временем пути. На рис. 12 есть точка M, из которой надо двигаться в точку N пешим способом, и точка M_1, из которой надо

двигаться в точку N_1 с помощью метро. Скажем, если человеку, живущему посредине между станциями «Рижская» и «Ботаническая», нужно добраться в район Земляного вала, то ему надо сесть на «Ботанической» и доехать до «Лермонтовской» или «Курской». Нетрудно видеть, что определенное таким образом расстояние $\varrho_t(M, N)$ существенно отличается от обычного геометрического расстояния. В самом деле, если точка Q расположена около вертолетной станции, т. е. между станциями «Динамо» и «Аэропорт», точка P в районе Волоколамского шоссе, а точка R в районе Валовой улицы (возле станции метро Павелецкая) (см. рис. 12), то в смысле обычного расстояния точка P гораздо ближе к Q, чем точка R,

$$\varrho(P, Q) < \varrho(Q, R).$$

Однако из карты на рис. 12 очевидно, что

$$\varrho_t(P, Q) > \varrho_t(P, R).$$

Действительно, если не пользоваться услугами такси, то от вертолетной станции до Валовой можно добраться быстрее, чем до Волоколамского шоссе.

Для расстояния $\varrho_t(M, N)$ нетривиальной является аксиома симметричности (свойство 1).

Равенство

$$\varrho_t(M, N) = \varrho_t(N, M)$$

означает, что время, затрачиваемое в обе стороны кратчайшего пути MN, одинаково. Это более или менее справедливо, если пользоваться метро и пешим способом сообщения. Но уже при пользовании такси это не так, ибо одно дело ждать такси недалеко от таксомоторного парка, а совсем другое — получить такси в каком-либо отдаленном районе, или у Курского вокзала в часы прихода поездов.

Свойства 2 и 3 для расстояния $\varrho_t(M, N)$ очевидны. Доказательство свойства 4 (аксиомы треугольника) читатель легко проведет самостоятельно, если вспомнит, как это свойство доказывалось в § 3 для расстояния $\varrho_{\text{П}}(M, N)$.

Для исследования свойств метрических пространств полезно ввести следующее понятие.

Пусть E — некоторое метрическое пространство, а L_1, L_2, ..., L_k — точки этого пространства.

Назовем *областью Дирихле точки* L_i множество D_i, состоящее из точек N, для которых

$$\varrho(L_i, N) \leqslant \varrho(L_j, N) \tag{28}$$

для всех $j \neq i$. Таким образом, область Дирихле D_i состоит из точек пространства, расположенных ближе к точке L_i, чем к остальным выбранным точкам. Ясно, что определение

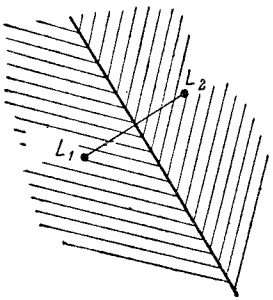

Рис. 13.

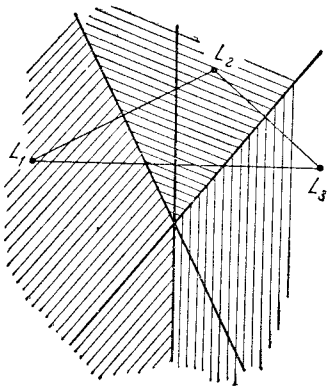

Рис. 14.

области Дирихле зависит от набора точек $L_1, L_2, \ldots, L_k$ и выбранной точки L_i. Рассмотрим примеры областей Дирихле в различных метрических пространствах.

Для начала возьмем плоскость с обычным расстоянием $\varrho_2(M, N)$ и две точки L_1 и L_2. Соединим эти точки отрезком L_1, L_2 (рис. 13) и проведем через середину этого отрезка перпендикуляр.

Этот перпендикуляр разбивает плоскость на две полуплоскости, которые и будут областями Дирихле для точек L_1 и L_2.

Возьмем теперь три точки L_1, L_2 и L_3 на плоскости с обычным расстоянием. На рис. 14 построены и отмечены штриховкой области Дирихле для этих трех точек. Способ построения ясен из рисунка.

Рассмотрим теперь две точки L_1 и L_2 на плоскости с расстоянием $\varrho_1(M, N)$ (т. е. в пространстве l). Для наглядности мы опять рассматриваем город, разделенный на кварталы. Области Дирихле состоят из тех перекрестков, от которых путь по городу будет короче, чем до точки L_1 (со-

ответственно L_2). Эти области отделены на рис. 15 жирной чертой. На рис. 16 приведено аналогичное разбиение для пространства C. Предоставляем читателю вывести общее правило построения областей Дирихле в пространствах l и C, для пар точек L_1 и L_2, а затем для троек L_1, L_2, L_3.

Обращаясь снова к рис. 12, мы видим, что если построить разбиение пространства на области Дирихле для пары точек P и R, то точка Q попадает в области Дирихле точки R.

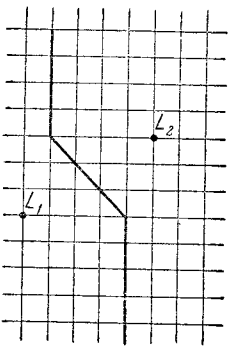

Рис. 15.

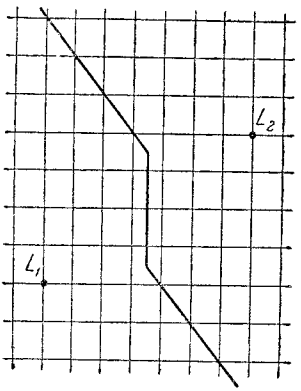

Рис. 16.

Предоставляем опять же читателю нарисовать это разбиение на области Дирихле. Важно, что это разбиение сильно отличается от разбиения на области Дирихле, показанного на рис. 13, 15 и 16.

§ 5. ПРОСТРАНСТВО СООБЩЕНИЙ

Под словом «сообщение» обычно понимается какая-то конкретная форма передачи информации. Сообщениями в широком смысле слова являются книги, письма, телеграммы, записи музыкальных произведений на пленку или с помощью нот, записи на перфокартах, с помощью которых информация о решаемой задаче вводится в математические машины, сигналы, управляющие взлетом и посадкой космических кораблей, молекулы дезоксирибонуклеиновой кислоты (ДНК), передающие наследственную информацию от родителей к потомству, и т. п.

Вопросы передачи и кодирования информации рассматривают в теории информации*). В теории информации разработаны способы определения количества информации, содержащейся в данном сообщении. Оказывается, что определенное количество информации может быть закодировано в виде сообщения разной величины. С этим обстоятельством мы часто встречаемся в жизни. Так, при составлении телеграммы мы стараемся сократить количество слов, не нарушая смысла (т. е. сохраняя количество информации).

Обратная ситуация возникает, когда плохо подготовленный студент на экзамене или семинаре начинает тянуть время, пытаясь небольшое количество имеющейся у него по данному вопросу информации выразить в виде достаточно внушительного количества слов.

Однако избыточная величина сообщения по сравнению с количеством передаваемой информации не всегда вредна. Эту избыточность можно обратить на пользу, когда при передаче сообщения возникают помехи.

Например, при плохой слышимости по телефону мы вынуждены повторять отдельные слова. При передаче незнакомого и трудного для восприятия названия или фамилии используется побуквенная система. Так, сообщая по телефону собственное имя «Павсикакий», мы будем произносить следующее: «Петр, Андрей, Василий, Семен, Илья, Константин, Анна, Кирилл, Инна, Иван краткий».

В этом и следующем параграфах мы, не вдаваясь в специфические вопросы теории информации, будем изучать некоторые приемы «помехоустойчивого кодирования» информации. Иными словами, мы будем изучать способы записи сообщений, позволяющие автоматически исправлять допущенные ошибки, если этих ошибок не слишком много. Эти приемы тесно связаны с возможностью определения расстояния в так называемом пространстве сообщений, которое изучается в настоящем параграфе.

Идею этих приемов все мы часто используем в обиходе, читая книги с опечатками и получая телеграммы с ошибками. Действительно, если в книге мы читаем слово «кистрюля», то не нужно заглядывать в перечень опечаток, чтобы догадаться о его значении. Очень мало шансов, что автор в этом

---

*) Основные сведения из этой теории в достаточно элементарном изложении можно найти в книге: А. М. Яглом и И. М. Яглом, Вероятность и информация, Физматгиз, 1960.

месте написал слово «телеграф». В самом деле, в этом случае мы имели бы дело с восемью опечатками подряд, в то время как из слова «кастрюля» тот же результат получается вследствие одной опечатки *).

Правда, существуют курьезные примеры, когда из-за ошибки в одной букве могут возникнуть различные слова. Скажем, слово «корона» может произойти и из «корова», и из «ворона».

На этом обстоятельстве основан известный анекдот, как одна провинциальная газета поместила в сообщении о коронации Николая Второго фразу: «митрополит возложил на голову Его Величества ворону». На следующий день было помещено исправление: «митрополит возложил на голову Его Величества корову».

Легко видеть, что и в этом случае истинный смысл сообщения довольно легко установить по контексту.

Аналогично ошибка в нотной записи часто может быть обнаружена по фальшивому звучанию и исправлена по законам гармонии.

Надо сказать, что точно так же ошибки могут возникать не только при передаче информации, но и при длительном ее хранении, например в памяти электронных вычислительных машин.

Проблема восстановления правильного сообщения имеет одинаковый вид и при передаче и при хранении информации.

Каждый вид сообщения записывается с помощью каких-то символов. Множество используемых символов образует алфавит $\mathfrak{A}$. Мы предположим, что этот алфавит задан заранее и состоит из конечного множества символов. Например, алфавит может состоять из всех русских букв, символа пропуска буквы и знаков препинания. В этом алфавите можно писать любые русские фразы. Другой пример алфавита — множество десятичных цифр, алгебраических знаков, знаков препинания и латинских и греческих букв. С помощью такого алфавита можно записывать разнообразнейшие математические формулы.

Еще один пример — это двоичный алфавит — множество из двух символов $\mathfrak{A}_2\{0, 1\}$. С помощью такого алфавита можно записывать числа в двоичной системе счисления.

---

*) Иногда бывают более вероятны групповые опечатки, связанные с неправильным пониманием машинисткой или наборщиком смысла данного слова.

Действительно, можно легко показать, что всякое целое число x записывается в виде

$$x = \varepsilon_n 2^n + \varepsilon_{n-1} 2^{n-1} + \ldots + \varepsilon_1 2 + \varepsilon_0, \qquad (29)$$

где величины ε_i принимают значения 0 или 1.

Таким образом, для того чтобы передать информацию о значении числа x, достаточно передать последовательность знаков алфавита $\mathfrak{A}_2$: $\varepsilon_n\ \varepsilon_{n-1}\ \ldots\ \varepsilon_1\ \varepsilon_0$. На самом деле, чтобы разделять сообщения о разных числах, надо либо ввести специальный знак для конца числа, либо передавать только последовательности двоичных знаков стандартной длины *).

Последний способ фактически применяется в вычислительных машинах, где запоминаемые двоичные последовательности обычно имеют стандартную величину, соответствующую величине так называемых ячеек памяти. В разрабатываемых в последние годы машинах от этого принципа постепенно отказываются, переходя к ячейкам переменной длины. Формула (29) аналогична хорошо известной формуле:

$$x = \alpha_n 10^n + \alpha_{n-1} 10^{n-1} + \ldots + \alpha_1 10 + \alpha_0, \qquad (30)$$

где α_n, α_{n-1}, ..., α_1, α_0 — цифры в десятичном представлении числа x. Легко распространить равенство (29) на не целые числа, подобно тому как вводятся десятичные дроби.

Установим связь между числом n и значением x в (29). Ясно, что если несколько первых слагаемых в правой части (29) равны нулю, то их можно отбросить. Поэтому мы будем предполагать, что $\varepsilon_n = 1$. Перенося все члены, кроме первого, из правой части в левую, мы получим

$$x - \varepsilon_{n-1} 2^{n-1} - \varepsilon_{n-2} 2^{n-2} \ldots - \varepsilon_1 2^1 - \varepsilon_0 = \varepsilon_n 2^n.$$

Отсюда ясно, что

$$2^n \leqslant x$$

или

$$n \leqslant \log_2 x. \qquad (31)$$

С другой стороны, верна следующая оценка:

$$\varepsilon_n 2^n + \varepsilon_{n-1} 2^{n-1} + \varepsilon_{n-2} 2^{n-2} + \ldots + \varepsilon_1 2^1 + \varepsilon_0 \leqslant$$
$$\leqslant 2^n + 2^{n-1} + 2^{n-2} + \ldots + 2^1 + 1 = 2^{n+1} - 1.$$

---

*) Есть и другие более сложные приемы разделения смысловых единиц (слов) в произвольном алфавите. См. статью А. А. Сардинаса и Дж. Паттерсона в «Кибернетическом сборнике», № 3, ИЛ, Москва, 1961.

Из последнего соотношения и из (29) следует

$$x \leqslant 2^{n+1} - 1$$

или

$$x < 2^{n+1},$$

что можно записать и так:

$$n + 1 > \log_2 x. \tag{32}$$

Сопоставляя (31) и (32), получаем двойное неравенство

$$n \leqslant \log_2 x \leqslant n + 1. \tag{33}$$

Неравенство (33) означает в других терминах, что

$$n = [\log_2 x],$$

т. е. n равно целой части *) $\log_2 x$ (читается: «n равно антье от $\log_2 x$»). Вышеизложенное можно записать еще так: количество двоичных знаков, необходимых для кодирования всех целых чисел в диапазоне $0 \leqslant x \leqslant a$, равно

$$1 + [\log_2 a] = 1 + n. \tag{34}$$

Единица здесь добавляется потому, что двоичные знаки включают еще и нулевой: ε_0.

С помощью двоичного алфавита $\mathfrak{A}_2$ в памяти вычислительных машин записывается любая информация (числа, команды, логические соотношения и др.) **).

Сообщением в данном алфавите $\mathfrak{A}$ мы будем называть конечную упорядоченную последовательность символов этого алфавита. Иногда удобно делить сообщения на некоторые стандартные сообщения, которые называются словами.

Вообще говоря, возможны также бесконечные алфавиты и сообщения, но мы их здесь рассматривать не будем.

Сообщение, записанное в одном алфавите, можно различными способами записать в другом алфавите.

Так, мы уже видели, что целое число, заданное десятичными цифрами, можно записать и в двоичном алфавите. Один из важных способов такой перекодировки состоит в следующем. Пусть дан алфавит $\mathfrak{A}$. Возьмем все слова, составленные в этом алфавите, длиной не более k и будем

---

*) *Целой частью числа* a называется наибольшее из целых чисел, меньших a. Целая часть a обозначается $[a]$; например, $[\pi] = 3$.

**) См. по этому вопросу: Б. А. Трахтенброт, Алгоритмы и машинное решение задач, изд. 2, Физматгиз, Москва, 1960).

считать это множество слов новым алфавитом $\mathfrak{A}'$. Ясно, что всякое сообщение в алфавите $\mathfrak{A}$ можно разбить на последовательность слов длины не более k, а значит, закодировать в новом алфавите $\mathfrak{A}'$.

Сходную идею можно было бы провести для сообщений на русском языке, записанных в русском алфавите, дополненном знаком пропуска и знаками препинания. Для этого нужно было бы взять полный словарь русского языка и каждое слово в нем обозначить иероглифом (используя, к примеру, комбинации китайских и египетских иероглифов). Если еще при этом ввести иероглифы так, что по ним можно различать падежи и спряжения глаголов, то можно будет закодировать любое сообщение на русском языке.

Вместо иероглифов можно было бы использовать шестизначные десятичные числа. Первых пяти знаков такого числа достаточно для кодирования слов*); шестой знак можно использовать для кодирования грамматических признаков.

Здесь мы впервые столкнулись с важным понятием кодирования и перекодирования сообщений. Под *кодированием* понимается, вообще говоря, формирование в заданном алфавите сообщений, несущих заданную информацию или преобразование сообщений, записанных в одном алфавите, в сообщения, записанные в другом алфавите. При этом наибольший интерес представляет взаимно однозначное перекодирование, когда по сообщению, записанному в любом из двух алфавитов, можно однозначно восстановить вид сообщения в другом алфавите. Нетрудно видеть, что рассмотренное перекодирование русских фраз из буквенного в «иероглифический» алфавит обладает указанным свойством.

Рассмотренный способ «пословного» кодирования практически используется при шифровке сообщений наряду с «побуквенной» шифровкой.

Возможна и обратная картина, когда символы исходного алфавита $\mathfrak{A}$ кодируются в виде слов, записанных в более простом алфавите $\mathfrak{A}'$. Например, пусть алфавит $\mathfrak{A}'$ состоит из трех символов $\{\cdot, —, *\}$; читается так: точка, тире, конец буквы. Тогда любая русская (или латинская) буква и знак препинания могут быть записаны в так называемой азбуке Морзе как слово из семи — не более — символов алфавита $\mathfrak{A}'$.

---

*) Так как заведомо можно удовлетвориться словарем из 100 000 русских слов.

Азбука Морзе *)

Символы Морзе	Русские буквы	Символы Морзе	Русские буквы
· —	а	· · — ·	ф
— · · ·	б	· · · ·	х
· — —	в	— · — ·	ц
— — ·	г	— — — ·	ч
— · ·	д	— — — —	ш
·	е	— — · —	щ
· · · —	ж	— · — —	ы
— — · ·	з	· · — —	ю
· ·	и	· — · —	я
— · —	к	· — — —	й
· — · ·	л	— · · —	ъ, ь
— —	м	· · — · ·	э
— ·	н	· — · · —	, (запятая)
— — —	о	· · · · · ·	. (точка)
· — — ·	п	— · — · — ·	; (точка с запятой)
· — ·	р	— — — · · ·	: (двоеточие)
· · ·	с	· · — — · ·	? (вопр. знак)
—	т	— — · · — —	! (воскл. знак)
· · —	у		

*) Знак конца буквы ✱ кодируется временным интервалом и поэтому не входит в приведенный в таблице код Морзе.

Таким образом, слова русского языка могут быть вместо русского алфавита записаны в алфавите Морзе.

Пример. Русская фраза «Что такое расстояние?» записывается в алфавите Морзе следующим образом:

— — — · ✱ — ✱ — — — ✱ — ✱ · — ✱ — · — ✱ — — — ✱ · ✱ · — · ✱ · — ✱ · · · ✱ · · · ✱ — ✱ — — — ✱ · — · — ✱ — · ✱ · · ✱ · ✱ · · — — · · ✱

Оказывается (и этим часто пользуются в вычислительной технике), сообщения в любом конечном алфавите могут быть перекодированы с помощью двоичного алфавита $\mathfrak{A}_2$ $\{0, 1\}$, состоящего только из двух символов «0» и «1».

Действительно, любое целое неотрицательное число может быть представлено в виде (29), т. е. в так называемой двоичной системе счисления. Так, каждое целое число x может быть закодировано в виде слова в двоичном алфавите.

Если целые числа меняются в диапазоне $0 \leqslant x < a$, то последовательность двоичных знаков $x \sim \varepsilon_n \varepsilon_{n-1} \varepsilon_{n-2} \ldots \varepsilon_1 \varepsilon_0$,

необходимая для записи любого из чисел этого диапазона, может быть выбрана из не более чем $n+1=1+[\log_2 a]$ знаков, как это следует из (34).

Теперь, если имеется произвольный алфавит $\mathfrak{A}$ из m символов, то каждому этому символу можно сопоставить порядковый номер от 0 до $m-1$. Значит, каждому символу алфавита $\mathfrak{A}$ можно сопоставить двоичное слово, соответствующее, согласно (29), номеру этого символа. При этом, согласно (24), можно обойтись словами длины n, где

$$\log_2(m-1) < n \leqslant 1+\log_2(m-1). \qquad (35)$$

Таким образом, в принципе можно было бы в любом случае ограничиться словами в двоичном алфавите. В современной автоматической телеграфии используется международный телеграфный код для записи русских и латинских букв, цифр и знаков препинания. Приведем для примера пятизначный код, используемый в телеграфных аппаратах СТ-35 *).

Латинский регистр	Числовой регистр	Русский регистр	Кодовая комбинация	Латинский регистр	Числовой регистр	Русский регистр	Кодовая комбинация
A	1	А	10000	U	4	У	10100
B	8	Б	00110	F	Э	Ф	01110
W		В	01101	H	+	Х	11010
G	7	Г	01010	C	9	Ц	10110
D	0	Д	11110	Q	/	Щ	10111
E	2	Е	01000	X	,	Ь	01001
V		Ж	11101	Y	3	Ы	00100
Z	:	З	11001			Я	00011
I	Ш	И	01100				
J	6	Й	10010				
K		К	10011	Русский регистр			11111
L	=	Л	11011				
M		М	01011	Числовой регистр			00010
N	Ю	Н	01111				
O	5	О	11100	Латинский регистр			00001
P		П	11000				
R	—	Р	00111	Пробел			10001
S	·	С	00101				
T	4	Т	10101	Звонок			00000

*) В настоящее время бо́льшее значение приобретает так называемый международный телеграфный код № 2, построенный по аналогичному принципу.

Последние пять комбинаций читаются во всех регистрах одинаково.

Знаки регистров означают, что после появления, скажем, знака латинского регистра все двоичные пятизначные комбинации читаются как латинские буквы. Чтобы перейти опять к русским буквам, нужно поставить знак русского регистра.

Пример. Запишем в рассмотренном телеграфном коде фразу: «Фамилия автора Гамлета по-английски пишется Shakespeare, а по-русски Шекспир»:

11111 01110 10000 01011 01100 11011 01100 00011 10001
10000 01101 10101 11100 00111 10000 10001 01010 10000
01011 11011 01000 10101 10000 10001 11000 11100 10001
10000 01111 01010 11011 01100 10010 00101 10011 01100
10001 11000 01100 **00010** 01100 **11111** 01000 10101 00101
00011 10001 **00001** 00101 11010 10000 10011 01000 00101
11000 01000 10000 00111 01000 10001 **00010** 01001 **11111**
10000 10001 11000 11100 10001 00111 10100 00101 00101
10011 01100 10001 **00010** 01100 **11111** 01000 10011 00101
11000 01100 00111 **00010** 00101

В кодированном тексте выделены знаки регистров, к которым приходится обращаться при смене русского и английского текстов, расстановке знаков препинания и написании буквы «ш». Последняя из-за большего объема русского алфавита по сравнению с латинским, помещена в числовом регистре.

Для записи только русских букв достаточно пятизначных двоичных символов без регистровых знаков. Соответствующая кодировка приведена в нижеследующей таблице.

а 00000	и 01000	р 10000	ш 11000
б 00001	й 01001	с 10001	щ 11001
в 00010	к 01010	т 10010	ъ 11010
г 00011	л 01011	у 10011	ы 11011
д 00100	м 01100	ф 10100	ь 11100
е 00101	н 01101	х 10101	э 11101
ж 00110	о 01110	ц 10110	ю 11110
з 00111	п 01111	ч 10111	я 11111

Фраза «Гипотенуза короче суммы катетов» запишется в этом коде так:

00011 01000 01111 01110 10010 00101 01101 10011 00111
00000 01010 01110 10000 01111 10111 00101 10001 10011
01100 01100 11011 01010 00000 10010 00101 10010 01110
00010

Заметим, что разделение пятизначных комбинаций мы ввели только для удобства чтения. При записи такого сообщения в машинной памяти или передаче по телеграфному каналу здесь не нужно никаких символов, кроме нуля и единицы.

В самом деле, представим себе, что вышенаписанный текст записан слитно (воспроизведем для примера первую строку):

$$0001101000011110111010010001010110110011.$$

Тогда мы можем сначала выделить первые пять знаков 00011 и написать их отдельно. Затем выделим следующие пять знаков 01000 и выпишем их отдельно. Действуя таким образом, мы получим сообщение с выделенными пробелами между словами (о разделимости слов в сообщении см. сноску на стр. 36).

Теперь введем понятие *пространства сообщений*. Рассмотрим произвольный алфавит $\mathfrak{A}$ *) и множество сообщений, состоящих ровно из n символов алфавита $\mathfrak{A}$.

Назовем *расстоянием между двумя сообщениями* ξ и η, $\varrho(\xi, \eta)$ количество позиций, в которых у сообщений ξ и η стоят различные символы. Полученное метрическое пространство $E(n, \mathfrak{A})$ называется n-мерным пространством сообщений в алфавите $\mathfrak{A}$.

Пример I. $\mathfrak{A}$ — русский алфавит, $n=6$. Пусть $\xi=$ корова; $\eta=$ ворона. Не совпадают первая и пятая буквы, следовательно, $\varrho(\xi, \eta)=2$.

Пример II. $\mathfrak{A}_2$ — двоичный алфавит, $n=12$ и $\xi=$ $=000110101010$; $\eta=010101101011$. Не совпадают 2-й, 5-й, 6-й и 12-й двоичные знаки, $\varrho(\xi, \eta)=4$.

Заметим, что можно было бы сравнивать любые слова длиной не более n, если условиться слова более короткие, чем n, дополнять определенным знаком (в двоичном алфавите нулем) до n символов.

*) В данном случае даже не существенно, что алфавит $\mathfrak{A}$ конечен.

Проверим, что определенное выше расстояние $\varrho(\xi, \eta)$ удовлетворяет всем необходимым свойствам.

Симметричность $\varrho(\xi \eta) = \varrho(\eta, \xi)$ следует из определения, в котором оба слова ξ и η участвуют равноправно. Очевидно, что $\varrho(\xi, \eta) \geqslant 0$, причем $\varrho(\xi, \eta) = 0$, только если все соответствующие символы в сообщениях ξ и η совпадают, т. е. совпадают слова ξ и η.

Аксиома треугольника проверяется так.

Пусть даны три слова ξ, η и ζ по n символов каждое. Предположим, что в k-й позиции совпадают символы у слов ξ и ζ и у слов ζ и η. Ясно, что тогда в этой позиции совпадают символы и у слов ξ и η.

Действительно, пусть ξ_k — k-й символ сообщения ξ, ζ_k — k-й символ сообщения ζ, а η_k — k-й символ сообщения η. Тогда, если $\xi_k = \zeta_k$, а $\zeta_k = \eta_k$, то, очевидно, $\xi_k = \eta_k$. Значит, если $\xi_k \neq \eta_k$, то либо $\xi_k \neq \zeta_k$, либо $\zeta_k \neq \eta_k$.

Таким образом, у слов ξ и η могут быть не совпадающие символы только на тех местах, где либо у ξ и ζ, либо у ζ и η не совпадают символы. Это значит, что общее количество не совпадающих символов у слов ξ и η не превосходит суммарного числа не совпадающих символов у ξ и ζ и у ζ и η. Но общее число символов, не совпадающих у ξ и ζ, плюс общее число символов, не совпадающих у ζ и η, есть $\varrho(\xi,\zeta) + \varrho(\zeta, \eta)$. Иными словами,

$$\varrho(\xi, \eta) \leqslant \varrho(\xi, \zeta) + \varrho(\zeta, \eta), \tag{36}$$

что и требовалось доказать.

Пример. В пространстве $E(6, \mathfrak{A})$, где $\mathfrak{A}$ — русский алфавит, запишем слова: $\xi =$ картон; $\eta =$ кортеж; $\zeta =$ кордон. Ясно, что $\varrho(\xi, \eta) = 3$; $\varrho(\xi, \zeta) = 2$, $\varrho(\zeta, \eta) = 3$ и

$$\varrho(\xi, \eta) \leqslant \varrho(\xi, \zeta) + \varrho(\zeta, \eta).$$

С помощью так определенного понятия расстояния можно сформулировать общий принцип построения кодов, позволяющих автоматически исправлять группы из не более чем l ошибок. Этот принцип впервые предложил Р. Хэмминг *). Он будет рассмотрен в следующем параграфе.

---

*) См. статью Р. Хэмминга в сб. переводов «Коды с обнаружением и исправлением ошибок», ИЛ, Москва, 1956.

§ 6. АВТОМАТИЧЕСКОЕ ИСПРАВЛЕНИЕ ОШИБОК В СООБЩЕНИЯХ

В этом параграфе мы рассматриваем пространство сообщений $E(n, \mathfrak{A})$, т. е. пространство сообщений длины n в алфавите $\mathfrak{A}$. Как мы видели ранее, фактически можно ограничиться двоичными сообщениями, т. е. сообщениями в алфавите $\mathfrak{A}_2$. Все содержательные примеры будут рассматриваться именно в этом алфавите.

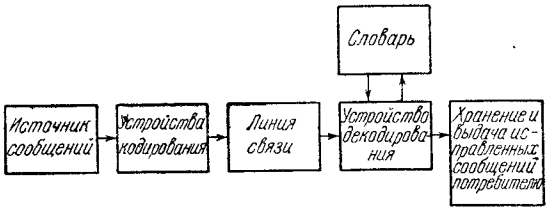

Рис. 17.

Рассмотрим следующую общую схему передачи информации (рис. 17). Сообщения, передаваемые источником, перекодируются в помехоустойчивый код устройством кодирования. Затем эти сообщения передаются по линиям

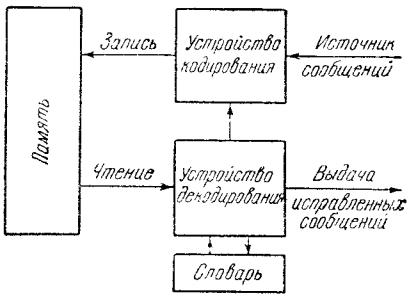

Рис. 18.

связи, в которых возможны искажения сообщений. После этого сообщения на приемном конце исправляются в декодирующем устройстве и снова перекодируются в исходный код, если это необходимо.

Совершенно аналогично происходит автоматическое обнаружение и исправление ошибок при хранении сообщений в машинной памяти (рис. 18).

При записи информации в память сообщения кодируются в помехоустойчивом коде. При чтении сообщения проходят соответствующую дешифровку с обнаружением и исправлением ошибок, допущенных при хранении. На рис. 18 видна дополнительная возможность, состоящая в том, что информация периодически считывается, проходит дешифровку с исправлением, а затем снова шифруется и записывается в память. Эта возможность бывает полезна, ибо она обеспечивает проверку хранимой информации не реже, чем один раз за T *сек*, где T надо выбрать так, чтобы за это время не могло возникнуть слишком много искажений в хранимой информации. Иначе говоря, чтобы расстояние $\varrho(\xi, \xi')$ между записанным сообщением ξ и считанным сообщением ξ' не успело стать слишком большим.

Пусть в пространстве $E(n, \mathfrak{A})$ выделено подмножество $H_k \subset E(n, \mathfrak{A})$, обладающее тем свойством, что для любых не совпадающих слов ξ и η множества H_k расстояние удовлетворяет условию

$$\varrho(\xi, \eta) \geqslant k. \tag{37}$$

Множество H_k назовем множеством осмысленных слов. Предположим, что при передаче или хранении осмысленного слова $\xi \in H_k$ допущено l ошибок ($l \leqslant k-1$), т. е. неверно передано l символов из слова ξ. Это ошибочно переданное слово обозначим ξ'. По определению расстояния $\varrho(\xi, \xi') = l$. Ясно, что слово ξ' не осмысленно, так как в противном случае $\varrho(\xi, \xi')$ было бы больше l (согласно (37)).

Таким образом, проверив переданное слово ξ' и убедившись, что оно не осмысленно (для этого можно его, к примеру, сравнить со всеми словами из H_k—на рис. 17 и 18 эта возможность обеспечивается наличием словаря), мы обнаруживаем ошибочность передачи. При хранении слов в машинной памяти такую проверку можно осуществлять периодически, выбирая период T из того условия, что за время T мало шансов для возникновения более чем $k-1$ ошибок в слове. Это рассуждение показывает, что у нас уже есть общий принцип для обнаружения ошибок.

Оказывается, можно сделать больше, а именно: исправить допущенные ошибки. Для этого предположим, что число ошибок $l \leqslant \frac{k-1}{2}$. Пусть η—произвольное осмысленное слово, не совпадающее с ξ, а ξ', как и ранее, ошибочно переданное сообщение.

Запишем неравенство треугольника

$$\varrho(\xi, \eta) \leqslant \varrho(\xi, \xi') + \varrho(\xi', \eta)$$

и подставим сюда $\varrho(\xi, \xi') = l$ и оценку (37):

$$k \leqslant l + \varrho(\xi', \eta).$$

Отсюда видно, что

$$\varrho(\xi', \eta) \geqslant k - l \geqslant k - \frac{k-1}{2} = \frac{k+1}{2}, \tag{38}$$

так как по условию $l \leqslant \frac{k-1}{2}$.

Из (38) видно, что ошибочное слово ξ' отстоит от любого осмысленного слова $\eta \neq \xi$ не менее чем на $\frac{k+1}{2}$, а от ξ не более чем на $\frac{k-1}{2}$. Следовательно, найдя ближайшее к ξ' осмысленное слово ξ, мы тем самым восстановим правильное сообщение.

Иными словами, нужно для осмысленных сообщений построить их области Дирихле.

Для каждого проверяемого (после передачи или хранения) сообщения ξ' определяется, чьей области Дирихле оно принадлежит. «Владелец» этой области Дирихле — осмысленное сообщение ξ — и считается правильным сообщением.

В этом и состоит замечательная идея Хэмминга. Она приводит к тому, что для передачи сообщений в условиях возможных искажений надо пользоваться не всеми возможными сочетаниями символов алфавита, а некоторым множеством осмысленных слов. Так как в русском языке используются в качестве осмысленных слов только некоторые сочетания букв, то восстановление смысла искаженного сообщения часто может быть выполнено без специальных дополнительных приемов кодирования. Это положение уже иллюстрировалось.

Мы сейчас рассмотрим несколько конкретных примеров построения множества осмысленных сообщений $H_k \in E(n, \mathfrak{A})$, т. е. выясним, как фактически, исходя из общего принципа Хэмминга, строить конкретные помехоустойчивые коды. Все эти примеры мы будем строить для случая двоичного алфавита $\mathfrak{A}_2$. Как мы уже видели, такое условие на самом деле не является ограничением, так как в двоичном алфавите можно записывать любые сообщения.

Задачу о помехоустойчивом кодировании можно поставить следующим образом. Пусть имеется пространство s-значных

двоичных сообщений $E(s, \mathfrak{A}_2)$. Требуется каждому такому сообщению поставить в соответствие сообщение из некоторого $H_k \subset E(n, \mathfrak{A}_2)$. Иными словами, нужно построить некоторое множество осмысленных сообщений H_k, устойчивое относительно l-кратных ошибок, с помощью которого можно кодировать исходные сообщения из $E(s, \mathfrak{A}_2)$. Величина $\frac{n-s}{n}$ называется *избыточностью* кода.

В действительности точная постановка этой задачи должна исходить из заданной статистики искажений в передающем канале. При этом нужно строить такой код (множество $H_k \subset E(n, \mathfrak{A}_2)$), что вероятность получить в слове из n двоичных символов более чем l ошибок будет достаточно мала. Такая постановка задачи изучается в теории информации, но мы ею здесь заниматься не будем.

Для построения этих кодов удобно использовать тот факт, что над величинами 0 и 1 можно выполнять особую операцию сложения по модулю 2 согласно правилам:

$$0 \oplus 0 = 0, \quad 1 \oplus 0 = 1,$$
$$0 \oplus 1 = 1, \quad 1 \oplus 1 = 0.$$

Знак операции сложения обведен кружком, чтобы показать, что эта операция несколько отличается от обычного сложения. Расстояние между двумя двоичными словами $\xi = x_1 x_2 \ldots x_n$ и $\eta = y_1 y_2 \ldots y_n$ (x_i и y_i принимают значения 0 или 1) можно с помощью этой операции записать так:

$$\varrho(\xi, \eta) = (x_1 \oplus y_1) + (x_2 \oplus y_2) + \ldots + (x_n \oplus y_n).$$

Действительно, в скобках здесь будут появляться единицы, когда соответствующие символы у обоих слов не совпадают, и нули в противном случае. Общее количество единиц как раз равно количеству не совпадающих символов в словах ξ и η.

Рассмотрим пространство сообщений $E(s, \mathfrak{A}_2)$ и сопоставим каждому слову $\xi \in E(s, \mathfrak{A}_2)$ слово ξ' длины $s+1$, образованное по следующему правилу.

Первые s символов слова ξ' совпадают со словом ξ.

Последний, $(s+1)$-й символ слова ξ' выбирается равным нулю или единице с тем, чтобы общее количество единиц в слове ξ' было четным. Если положить $\xi' = x_1 x_2 \ldots x_s x_{s+1}$, то это условие означает, что

$$x_1 \oplus x_2 \oplus \ldots x_s \oplus x_{s+1} = 0. \tag{39}$$

Отсюда можно выразить x_{s+1}:

$$x_{s+1} = x_1 \oplus x_2 \oplus \ldots \oplus x_s. \qquad (40)$$

Например, если $\xi = 001\,011\,101$, то $\xi' = 001\,011\,101\,1$.

Определенные таким образом слова ξ' образуют множество осмысленных сообщений $H_2 \subset E(s+1, \mathfrak{A}_2)$.

В самом деле, расстояние между двумя не совпадающими осмысленными сообщениями ξ' и ξ'' должно быть четным, так как, если бы ξ' отличалось от ξ'' в нечетном числе позиций, то сумма единиц, хотя бы в одном из слов ξ' или ξ'', была бы нечетной. А это невозможно по построению этих слов. Так как минимальное четное число, не равное нулю, есть 2, то минимальное расстояние между различными осмысленными сообщениями равно двум.

Следовательно, этот код позволяет обнаруживать единичные ошибки. Процесс обнаружения состоит просто в том, что после передачи слова ξ' подсчитывается четность единиц в нем (или, что равносильно, проверятся (39)). Если критерий четности не удовлетворяется ((39) не выполнено), то принятое слово считается ошибочным. Этот процесс проверки широко известен как «проверка на четность» и очень часто применяется ввиду его простоты. Избыточность в этом случае равна $\frac{1}{s+1}$.

Опишем теперь принадлежащий Хэммингу красивый пример кода, способного исправлять единичные ошибки.

Пусть $\xi \in E(s, \mathfrak{A}_2)$ — двоичное слово длины s. Образуем слово $\xi' \in E(n, \mathfrak{A}_2)$ по следующему правилу. Среди n-позиций, которые занимает слово ξ', мы выделим 1-ю, 2-ю, 4-ю, ..., 2^k-ю позицию для контрольных символов, которые определяются по слову ξ. Между этими позициями мы будем последовательно записывать символы, образующие слово:

$$\xi = \mathbf{10011100}1001101111111011000\mathbf{10}10.$$

В написанном слове показано взаимное расположение контрольных (выделенных жирным шрифтом) и смысловых позиций для случая $s=25$, $n=31$, $k=4$. Для того чтобы разместить нужным образом s смысловых позиций, необходимо и достаточно, чтобы выполнялось соотношение

$$2^k \leqslant s + k < 2^{k+1}. \qquad (41)$$

Избыточность данного кода равна $\frac{k}{s+k}$.

Контрольная позиция с номером 2^i заполняется по следующему правилу. Каждая позиция слова ξ' определяется номером l, отсчитываемым от начала слова. Рассмотрим двоичное представление этого номера:

$$l = l_k 2^k + l_{k-1} 2^{k-1} + \ldots + l_1 2 + l_0$$

(число двоичных разрядов в представлении номера l определяется тем, что согласно (41) $l < 2^{k+1}$).

Возьмем теперь множество Π_i всех тех позиций, для номеров которых $l_i = 1$. В множество Π_i входит одна контрольная позиция, а именно позиция с номером $l = 2^i$. Заполним ее так, чтобы сумма всех единиц в позициях из Π_i была четной.

В следующей таблице приведен пример слова ξ' (его можно прочесть по вертикали во втором столбце таблицы), указаны двоичные номера позиций и отмечена знаком * принадлежность каждой позиции к множеству Π_i. Слова ξ', построенные по этому правилу, назовем осмысленными.

№ позиции	Содержимое позиции	Π_0	Π_1	Π_2	Π_3	Π_4
00 001		*				
00 010			*			
00 011	0	*	*			
00 100				*		
00 101	1	*		*		
00 110	1		*	*		
00 111	0	*	*	*		
01 000					*	
01 001	1	*			*	
01 010	0		*		*	
01 011	0	*	*		*	
01 100	1			*	*	
01 101	1	*		*	*	
01 110	0		*	*	*	
01 111	1	*	*	*	*	
10 000						*
10 001	1	*				*
10 010	1		*			*
10 011	1	*	*			*
10 100	1			*		*
10 101	0	*		*		*
10 110	1		*	*		*
10 111	1	*	*	*		*
11 000	0				*	*

Продолжение

№ позиции	Содержимое позиции	Π_0	Π_1	Π_2	Π_3	Π_4
11 001	0	*			*	*
11 010	0		*		*	*
11 011	1	*	*		*	*
11 100	0			*	*	*
11 101	0	*		*	*	*
11 110	1		*	*	*	*
11 111	0	*	*	*	*	*

Покажем, что расстояние между любыми двумя осмысленными словами ξ' и η' не менее чем 3, то есть осмысленные слова образуют множество $H_s \subset E(s+k, \mathfrak{A}_2)$.

I случай. Пусть соответствующие слова ξ и η из $E(s, \mathfrak{A}_2)$ отличаются не менее чем в трех позициях. Ясно, что слова ξ' и η' также отличаются не менее чем в трех позициях, а следовательно, $\varrho(\xi', \eta') \geqslant 3$.

II случай. Пусть слова ξ и η отличаются в двух позициях. Тем самым слова ξ' и η' отличаются в двух смысловых позициях с номерами l и l'. Выберем множество позиций Π_i, содержащее позицию l и не содержащее позиции l'. Для этого нужно номер i выбрать так, чтобы i-й двоичный знак числа l, $l_i=1$, а i-й двоичный знак числа l', $l'_i=0$ (так как $l \neq l'$, то у них обязательно есть не совпадающие двоичные знаки).

Таким образом, у слов ξ' и η' количество единиц в смысловых позициях множества Π_i отличается на одну. Так как общее число единиц во множестве позиций Π_i должно быть четным и у ξ', и у η', то слова ξ' и η' должны отличаться еще и в контрольной позиции множества Π_i (в позиции с номером 2^i). Тем самым ξ' и η' отличаются не менее чем в трех позициях, $\varrho(\xi', \eta') \geqslant 3$.

III случай. Пусть слова ξ и η отличаются в одной позиции. Тогда слова ξ' и η' отличаются ровно в одной смысловой позиции с номером l. Этот номер не может быть точной степенью двойки, так как номера вида 2^i закреплены за контрольными позициями. Таким образом, у номера l есть по крайней мере два ненулевых двоичных знака $l_i=1$ и $l_j=1$. Следовательно, позиция с номером l входит в два множества Π_i и Π_j. Так как количество единиц в этих множествах

позиций и для ξ', и для η' должно быть четным, то кроме l-й позиции слова ξ' и η' обязательно должны отличаться и в контрольных позициях с номерами 2^i и 2^j.

Всего, таким образом, слова ξ' и η' отличаются не менее чем в трех позициях, $\varrho(\xi', \eta') \geqslant 3$.

Итак, доказано, что множество осмысленных сообщений есть H_3 и, следовательно, позволяет принципиально восстанавливать единичные ошибки.

В данном случае этот процесс восстановления может быть сравнительно просто выполнен. В основе этого лежит такое соображение. Для исправления слова в двоичном алфавите достаточно узнать номер позиции, где допущена ошибка и переменить содержимое в этой позиции (если был «0», поставить «1» и наоборот).

В рассматриваемом коде номер неисправной позиции устанавливается путем следующих проверок.

После передачи сообщения ξ', в результате которой возможно искажение в одной позиции $\xi' \longrightarrow \xi^*$, проверяется четность единиц для каждого множества позиций Π_i.

Иными словами, вычисляются контрольные величины

$$\alpha_i = \xi_1^* \oplus \xi_2^* \oplus \xi_3^* \oplus \ldots,$$

т. е. α_i равна сумме по модулю 2 всех символов в позициях множества Π_i принятого сообщения ξ^*.

Если все $\alpha_i = 0$, то ξ^* является осмысленным сообщением*). Если же некоторое $\alpha_i = 1$, то ошибка допущена в позиции, принадлежащей множеству Π_i, а значит, номер l этой позиции имеет i-й двоичный знак, равный 1, $l_i = 1$.

Таким образом, контрольные величины α_i равны знакам двоичного разложения номера ошибочной позиции l, т. е.

$$l = \alpha_k 2^k + \alpha_{k-1} 2^{k-1} + \ldots + \alpha_1 2 + \alpha_0. \qquad (42)$$

Это значит, что после всех контрольных вычислений номер l определен и дает возможность восстановить правильное сообщение ξ'.

Для примера возьмем слово ξ', записанное в предыдущей таблице, и исказим его в 19-й позиции.

Получится слово

$$\xi^* = 1001110010011011010\ 101100010010;$$

*) То есть если и была допущена ошибка, то не менее чем в трех позициях, и ее нельзя исправить с помощью данного кода.

осуществляя все проверки, получаем

$$\alpha_4 = 1,\ \alpha_3 = 0,\ \alpha_2 = 1,\ \alpha_1 = 0,\ \alpha_0 = 1,$$

т. е. $l = 10\,101$, что соответствует десятичному числу 19. Переменив в 19 позиции слова ξ^* 0 на 1, получаем передававшееся слово ξ'.

Более простой код с исправлением одиночных ошибок будет, если слово ξ' получается трехкратным повторением слова $\xi \in E(s, \mathfrak{A}_2)$. Если ξ и η отличаются в r позициях, то соответствующие ξ' и η' отличаются в $3r$ позициях. Таким образом, $\varrho(\xi', \eta') \geqslant 3$, если $\xi' \neq \eta'$. Проверка переданного слова осуществляется так.

Берутся тройки позиций с номерами l, $l+s$, $l+2s$. Если символы в этих позициях совпадают, то соответствующий символ считается правильно переданным. Если совпадают два из этих символов, то их общее значение считается правильным и заносится в третью позицию.

Если все три символа не совпадают, то ошибку исправить нельзя.

Таким образом, этот код может исправлять единичные ошибки в каждой тройке соответствующих позиций и обнаруживать двойные ошибки в каждой тройке.

Недостатком этого кода является его большая избыточность, равная $\frac{2s}{3s} = \frac{2}{3}$. Избыточность предыдущего кода примерно равна $\frac{\log_2 s}{s}$, что при больших s (длинных сообщениях) близко к нулю.

Коды, позволяющие исправлять ошибки при передаче и хранении сообщений, очень важны для различных устройств автоматического управления. За последние 10 лет появилось большое количество работ, где изучены коды, исправляющие единичные или групповые ошибки.

§ 7. РАССТОЯНИЯ И НОРМЫ В МНОГОМЕРНОМ ПРОСТРАНСТВЕ

В этом параграфе рассматривается так называемое n-мерное векторное пространство E_n и различные определения расстояния в нем, превращающие E_n в различные метрические пространства. Векторное пространство E_n служит обобщением понятий прямой, плоскости и трехмерного пространства из элементарной геометрии. Прийти к разумному

определению n-мерного пространства можно следующим образом.

Рассмотрим плоскость с выбранной системой декартовых координат. Каждая точка M на плоскости однозначно определяется своими координатами (x, y).

Каждой точке M однозначно соответствует вектор (см. рис. 19), соединяющий начало координат с этой точкой. Таким образом, существует взаимное однозначное соответствие между следующими объектами:

$$\text{точка } M \leftrightarrow \text{вектор } \overrightarrow{OM} \leftrightarrow \text{пара координат } (x, y).$$

Это значит, что можно с одинаковым успехом говорить о плоскости как о множестве точек, как о множестве векторов или как о множестве пар чисел (x, y) *). Аналогично, о трехмерном пространстве можно говорить как о множестве троек чисел (x, y, z). Теперь уже сами собой напрашиваются такие определения.

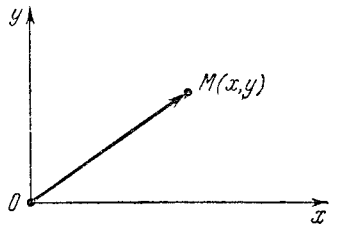

Рис. 19.

n-мерным вектором называется упорядоченная совокупность из n вещественных чисел

$$\xi = (x_1, x_2, \ldots, x_n).$$

Числа $x_1, x_2, \ldots, x_n$ называются *координатами вектора* ξ. Множество всех n-мерных векторов называется *n-мерным векторным пространством* E_n.

Ясно, что векторное пространство E_3 есть обычное трехмерное пространство, пространство E_2 есть плоскость, а пространство E_1 — прямая.

Над векторами в n-мерном пространстве можно выполнять операции сложения, вычитания и умножения на вещественное число. Эти операции определяются следующим образом.

Суммой векторов $\xi = (x_1, x_2, \ldots, x_n)$ и $\eta = (y_1, y_2, \ldots, y_n)$ называется вектор

$$\zeta = \xi + \eta = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n),$$

*) То есть говорить об одном и том же математическом объекте — плоскости — на языке точек, на языке векторов или на языке числовых пар. Любое утверждение может быть переведено с каждого языка на каждый.

у которого координаты равны суммам соответствующих координат слагаемых.

Аналогично, *разностью тех же векторов* ξ и η называется вектор

$$\theta = \xi - \eta = (x_1 - y_1,\ x_2 - y_2, \ldots,\ x_n - y_n),$$

у которого координаты равны разностям соответствующих координат.

Произведением вектора $\xi = (x_1,\ x_2, \ldots,\ x_n)$ на число a называется вектор

$$\varphi = a\,\xi = (ax_1,\ ax_2,\ \ldots,\ ax_n).$$

Иными словами, чтобы умножить вектор на число, нужно умножить на это число все его координаты. На плоскости

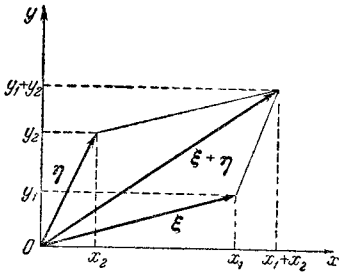

Рис. 20.

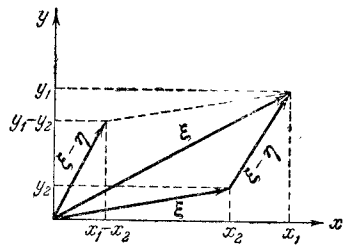

Рис. 21.

и в трехмерном пространстве эти операции имеют простой геометрический смысл. Для наглядности рассмотрим два вектора ξ и η на плоскости (см. рис. 20). Из рисунка видно, что сумма $\zeta = \xi + \eta$ есть вектор, образованный диагональю параллелограмма, построенного на векторах ξ и η. Такое определение суммы векторов возникает в физике при сложении сил.

Разность векторов ξ и η (см. рис. 21) есть вектор, направленный из конца вектора η в конец вектора ξ.

Произведение вектора ξ на положительное число a есть вектор, имеющий то же направление, а длину, в a раз большую, чем ξ. (Ясно, что при $a < 1$ длина вектора $a\xi$ меньше длины вектора ξ.) Чтобы умножить вектор ξ на отрица-

тельное число a, надо умножить его на $|a|$, а затем изменить направление на противоположное. Все эти случаи показаны на рис. 22.

Легко показать, что операции над n-мерными векторами удовлетворяют следующим свойствам, аналогичным свойствам соответствующих числовых операций.

1. $\xi+\eta=\eta+\xi$ (коммутативность),
2. $\xi+(\eta+\zeta)=(\xi+\eta)+\zeta$ (ассоциативность),
3. $\xi-\xi=0$,
4. $0+\xi=\xi$,
5. $\lambda(\xi+\eta)=\lambda\xi+\lambda\eta$ (дистрибутивность),
6. $(\lambda+\mu)\xi=\lambda\xi+\mu\xi$ (дистрибутивность),
7. $\lambda(\mu\xi)=(\lambda\mu)\xi$ (ассоциативность умножения),
8. $0\cdot\xi=0$,
9. $\lambda\cdot 0=0$,
10. $1\cdot\xi=\xi$.

Здесь использовано обозначение для нулевого вектора $0=(0, 0, \ldots, 0)$. Ясно, что два вектора ξ и η совпадают в том и только том случае, когда $\xi-\eta=0$.

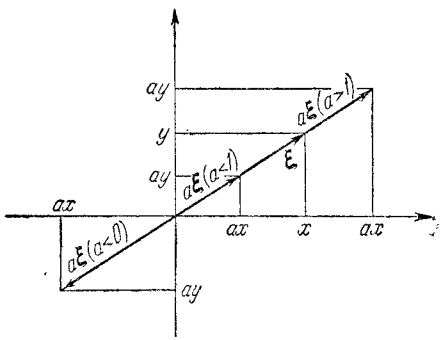

Рис. 22.

Рассмотрим несколько примеров многомерных пространств, естественно возникающих в геометрии.

Пример I. Множество всех шаров в пространстве. Каждый шар определяется четырьмя параметрами (x, y, z, R), где (x, y, z) — координаты центра, а R — радиус.

Пример II. Множество всех треугольников. Каждый треугольник определяется девятью параметрами $(x_1, y_1, z_1, x_2, y_2, z_2, x_3, y_3, z_3)$, где каждая тройка (x_i, y_i, z_i) состоит из координат некоторой вершины. Предоставляем читателю убедиться, что в обоих случаях умножение всех координат (соответственно четырех- и девятимерного вектора) на число λ равносильно подобному преобразованию с центром подобия в начале координат. Более детальное изучение пространств шаров и треугольников с различными вариантами расстояния также предоставляем читателю.

Теперь рассмотрим различные способы определения расстояния в E_n, превращающие E_n в различные метрические пространства.

Пространство $l_2^{(n)}$ получается, если расстояние определить формулой

$$\varrho_2(\xi, \eta) = \sqrt{(x_1-y_1)^2+(x_2-y_2)^2+\ldots+(x_n-y_n)^2}, \quad (43)$$

где

$$\xi = (x_1, x_2, \ldots, x_n), \text{ а } \eta = (y_1, y_2, \ldots, y_n).$$

В трехмерном пространстве и на плоскости расстояние $\varrho_2(\xi, \eta)$ совпадает с обычным геометрическим расстоянием. Свойства 1, 2, 3 для этого расстояния очевидны.

Пространство $l_1^{(n)}$ определяется с помощью расстояния

$$\varrho_1(\xi, \eta) = |x_1-y_1| + |x_2-y_2| + \ldots + |x_n-y_n|. \quad (44)$$

В случае плоскости это расстояние совпадает с расстоянием $\varrho_1(M, N)$, определенным в § 4. Опять-таки, свойства 1, 2 и 3 для расстояния $\varrho_1(\xi, \eta)$ очевидны.

Пространство $C^{(n)}$ возникает, если ввести расстояние по правилу

$$\varrho_\infty(\xi, \eta) = \max(|x_1-y_1|, |x_2-y_2|, \ldots |x_n-y_n|), \quad (45)$$

т. е. как максимальное отклонение координат соответствующих векторов. Свойства 1, 2 и 3 очевидны и для этого расстояния, которое на плоскости совпадает с введенным в § 4 расстоянием $\varrho_\infty(M, N)$.

Докажем аксиому треугольника для пространства $l_1^{(n)}$.

Пусть

$$\xi = (x_1, x_2, \ldots, x_n); \ \eta = (y_1, y_2, \ldots, y_n); \ \zeta = (z_1, z_2, \ldots, z_n).$$

Тогда, очевидно,

$$\begin{aligned}\varrho_1(\xi, \eta) &= |x_1-y_1|+|x_2-y_2|+\ldots+|x_n-y_n| = \\ &= |x_1-z_1+z_1-y_1|+|x_2-z_2+z_2-y_2|+\ldots+ \\ &+|x_n-z_n+z_n-y_n| \leqslant |x_1-z_1|+|z_1-y_1|+|x_2-z_2|+ \\ &+|z_2-y_2|+\ldots+|x_n-z_n|+|z_n-y_n| = \\ &= \varrho_1(\xi, \zeta)+\varrho_1(\zeta, \eta),\end{aligned}$$

что и требовалось доказать.

Для пространства $C^{(n)}$ аксиома треугольника доказывается так. Пусть $|x_k-y_k|$ самая большая из соответствующих разностей координат, т. е.

$$\begin{aligned}\varrho_\infty(\xi, \eta) &= \max(|x_1-y_1|,\ |x_2-y_2|,\ \ldots,\ |x_n-y_n|) = \\ &= |x_k-y_k| = |x_k-z_k+z_k-y_k| \leqslant \\ &\leqslant |x_k-z_k|+|z_k-y_k|. \qquad (46)\end{aligned}$$

Очевидно, что

$$\left.\begin{aligned}|x_k-z_k| &\leqslant \max(|x_1-z_1|,\ |x_2-z_2|,\ \ldots,\ |x_n-z_n|) = \\ &= \varrho_\infty(\xi, \zeta), \\ |z_k-y_k| &\leqslant \max(|z_1-y_1|,\ |z_2-y_2|,\ \ldots,\ |z_n-y_n|) = \\ &= \varrho_\infty(\zeta, \eta).\end{aligned}\right\} \qquad (47)$$

Сравнивая (46) и (47), получаем искомое соотношение

$$\varrho_\infty(\xi, \eta) \leqslant \varrho_\infty(\xi, \zeta)+\varrho_\infty(\zeta, \eta).$$

Более общий вид метрических пространств получится, если ввести расстояние по формуле

$$\varrho_p(\xi, \eta) = \sqrt[p]{(x_1-y_1)^p+(x_2-y_2)^p+\ldots+(x_n-y_n)^p},$$

где $p \geqslant 1$, и получить таким образом пространство $l_p^{(n)}$.

Свойства 1, 2 и 3 проверяются здесь столь же легко. Неравенство треугольника (свойство 4) получается из неравенства Минковского (см. ссылку на стр. 28):

$$\begin{aligned}&\sqrt[p]{(a_1-b_1)^p+(a_2-b_2)^p+\ldots+(a_n-b_n)^p} \leqslant \\ &\leqslant \sqrt[p]{a_1^p+a_2^p+\ldots+a_n^p}+\sqrt[p]{b_1^p+b_2^p+\ldots+b_n^p}.\end{aligned}$$

Нетрудно видеть, что при $p=1$ и $p=2$ получаются выше определенные пространства $l_1^{(n)}$ и $l_2^{(n)}$. При $p \to \infty$ расстояние $\varrho_p(\xi, \eta)$ переходит в расстояние $\varrho_\infty(\xi, \eta)$, т. е. $l_\infty^{(n)} = C^{(n)}$. Этот факт читатель сумеет легко проверить, обобщив аналогичные рассуждения из § 4.

В качестве более трудного упражнения читателю предлагается показать, что обобщенный шар радиуса r в пространстве $l_1^{(3)}$ имеет вид октаэдра (рис. 23), а в пространстве $C^{(3)}$ имеет вид куба (рис. 24).

Пространства $l_p^{(n)}$, так же как и соответствующие пространства l_p на плоскости, называются пространствами Минковского. Эти пространства допускают красивые обобщения на случай бесконечного количества координат у вектора.

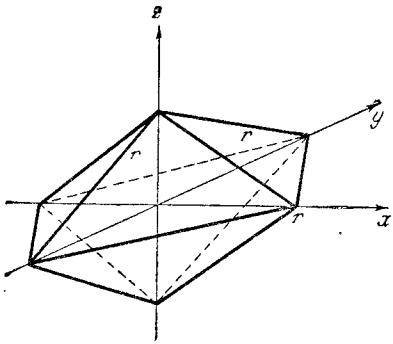

Рис. 23.

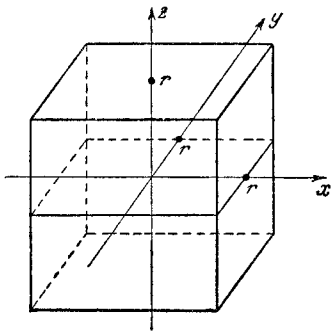

Рис. 24.

Более общий вид расстояния можно ввести с помощью понятия выпуклого тела.

Введем несколько новых определений. Каждый вектор из векторного пространства E_n мы будем геометрически интерпретировать как точку, соответствующую концу этого вектора, считая сам вектор выходящим из начала координат. Поэтому каждому факту, относящемуся к векторам, соответствует некоторое свойство точек n-мерного пространства. Назовем подмножество V пространства E_n *выпуклым*, если вместе с каждыми двумя векторами $\xi \in V$ и $\eta \in V$ оно содержит любой вектор вида $a\xi + (1-a)\eta$, где a — любое число между нулем и единицей ($0 \leqslant a \leqslant 1$). Геометрически (в E_2 и в E_3) это означает, что множество V вместе с любой парой точек содержит и весь отрезок, соединяющий эти точки.

Назовем подмножество $V \subset E_n$ *ограниченным*, если существует такое общее для всех векторов число K, что координаты всех векторов $\xi \in V$ удовлетворяют неравенствам

$$|x_1| < K;\ |x_2| < K;\ \ldots\ |x_n| < K.$$

На рис. 25 изображено выпуклое, но не ограниченное множество на плоскости. А на рис. 26 — выпуклое и ограниченное множество.

Вектор ξ, принадлежащий множеству V, называется *внутренним* для множества V, если для всякого вектора η можно подобрать такое положительное число a, что вектор $\xi + a\eta$ также входит в множество V. Это значит, что если из конца вектора ξ двигаться по любому направлению, то некоторое время мы будем еще находиться во множестве V.

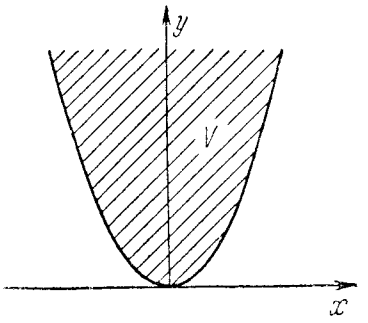

Рис. 25.

Если V — плоская фигура, расположенная в пространстве E_3, то никакая его точка не является внутренней. Действительно (см. рис. 27), если $\xi \in V$, а вектор η перпендикулярен плоскости, где лежит V,

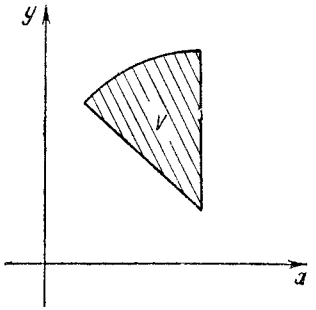

Рис. 26.

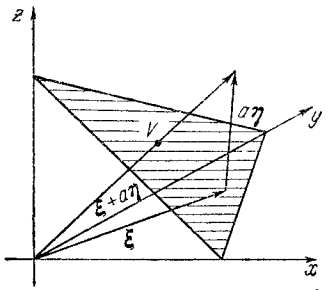

Рис. 27.

то при любом a вектор $\xi + a\eta$ лежит вне этой плоскости, а следовательно, и вне фигуры V.

Наличие во множестве V внутренних точек делает его «*полноразмерным*».

Назовем, наконец, *выпуклым телом* выпуклое ограниченное множество V, имеющее хотя бы одну внутреннюю точку *).

*) Читателю предоставляется доказать, что в этом случае множество V обязательно имеет бесконечно много внутренних точек.

Рассмотрим выпуклое тело V, имеющее O своей внутренней точкой и симметричное относительно этой точки. Последнее означает, что вместе с каждым вектором ξ в теле V содержится и противоположный вектор $-\xi$.

С помощью этого выпуклого тела можно ввести расстояние $\varrho_V(\xi, \eta)$. Для этого используем такую конструкцию. Образуем разность векторов $\zeta = \xi - \eta$. Так как O является внутренней очкой тела V, то существует такое положительное число a, при котором $a\zeta \in V$. С другой стороны, из ограниченности тела V следует, что при достаточно больших a вектор $a\zeta$ лежит вне тела V, если сам вектор $\zeta \neq 0$. Определим расстояние $\varrho_V(\xi, \eta)$ как нижнюю грань величин $\frac{1}{a}$, $a > 0$, для которых вектор $a\zeta$ еще принадлежит телу V,

$$\varrho_V(\xi, \eta) = \inf_{a\zeta \in V} \frac{1}{a}. \tag{48}$$

При $\zeta = 0$, т. е. когда векторы ξ и η совпадают, допустимо любое a (так как всегда $a\zeta = 0 \in V$) и нижняя грань в (48) равна нулю. При несовпадающих ξ и η вектор $\zeta \neq 0$ и допустимые значения a ограничены сверху. С другой стороны, при достаточно малых $a > 0$, $a\zeta \in V$; поэтому значения величины $\frac{1}{a}$ также ограничены сверху. Следовательно, в этом случае величина $\varrho_V(\xi, \eta)$ конечна, положительна и не равна нулю *).

Из симметричности выпуклого тела V следует, что векторы $a\zeta$ и $a(-\zeta)$ принадлежат телу V при одних и тех же значениях a. Так как $-\zeta = \eta - \xi$, то отсюда следует симметричность расстояния

$$\varrho_V(\xi, \eta) = \varrho_V(\eta, \xi).$$

Итак, для определенного нами с помощью выпуклого тела V расстояния доказаны свойства 1, 2 и 3 (см. стр. 18). Осталось доказать неравенство треугольника.

Это можно сделать следующим образом.

Рассмотрим три вектора ξ, η и ζ. Выберем два положительных числа a и b такие, что $a(\xi - \zeta) \in V$ и $b(\zeta - \eta) \in V$.

*) Здесь мы пользуемся свойствами нижней грани, которые неискушенному читателю очевидны, а знакомому с понятием нижней грани известны.

Обозначим теперь через α число

$$\alpha = \frac{\frac{1}{a}}{\frac{1}{a}+\frac{1}{b}} = \frac{b}{a+b}.$$

Ясно, что $0 < \alpha < 1$ и $1-\alpha = \frac{a}{a+b}$. В силу выпуклости тела V вектор

$$0 = \alpha[a(\xi-\zeta)] + (1-\alpha)[b(\zeta-\eta)] \tag{49}$$

также содержится в V. Преобразуем теперь выражение (49), подставляя вместо α и $1-\alpha$ их значения и пользуясь свойствами операций над векторами:

$$0 = \frac{b}{a+b}\cdot a(\xi-\zeta) + \frac{a}{a+b}\cdot b(\zeta-\eta) =$$

$$= \frac{ab}{a+b}[\xi-\zeta+\zeta-\eta] = \frac{ab}{a+b}(\xi-\eta) = c(\xi-\eta). \tag{50}$$

Так как вектор $0 = c(\xi-\eta) \in V$, то число

$$\frac{1}{c} = \frac{a+b}{ab} = \frac{1}{b}+\frac{1}{a}$$

может быть только больше расстояния $\varrho_V(\xi, \eta)$:

$$\varrho_V(\xi, \eta) \leqslant \frac{1}{a}+\frac{1}{b}. \tag{51}$$

Однако в силу определения чисел a и b величины $\frac{1}{a}$ и $\frac{1}{b}$ можно сделать сколь угодно близкими, соответственно к расстояниям $\varrho_V(\xi, \zeta)$ и $\varrho_V(\zeta, \eta)$. Так как при переходе к пределу знак неравенства сохраняется, то из (51) получается искомое неравенство

$$\varrho_V(\xi, \eta) \leqslant \varrho_V(\xi, \zeta) + \varrho_V(\zeta, \eta), \tag{52}$$

справедливое для любых троек векторов.

Определение расстояния с помощью симметричного выпуклого тела V можно изложить в несколько иных терминах.

Назовем *нормой вектора* ξ величину

$$\|\xi\|_V = \inf_{a\xi \in V} \frac{1}{a}. \tag{53}$$

Ясно, что определенное выше расстояние $\varrho_V(\xi, \eta)$ равно норме разности векторов ξ и η

$$\varrho_V(\xi, \eta) = \|\xi-\eta\|_V. \tag{54}$$

Можно показать (оставляем доказательство читателю), что норма удовлетворяет следующим свойствам:

1. $\|\xi\| \geqslant 0$;
2. $\|\xi\|_V = 0$ тогда и только тогда, когда $\xi = \mathbf{0}$;
3. $\|a\xi\|_V = |a| \|\xi\|_V$, если a — любое вещественное число;
4. $\|\xi + \eta\|_V \leqslant \|\xi\|_V + \|\eta\|_V$.

Можно было бы подойти к понятию нормы иным, более абстрактным способом. Назовем *нормой* в n-мерном векторном пространстве E_n число $\|\xi\|$, поставленное в соответствие каждому вектору ξ, так что при этом выполнены следующие свойства:

1. $\|\xi\| \geqslant 0$;
2. $\|\xi\| = 0$ тогда и только тогда, когда $\xi = 0$;
3. $\|a\xi\| = |a| \|\xi\|$;
4. $\|\xi + \eta\| \leqslant \|\xi\| + \|\eta\|$.

Векторное пространство E_n с определенной в нем нормой называется *n-мерным пространством Минковского* *). Оказывается, что норма всегда может быть определена с помощью некоторого симметричного выпуклого тела V. Проверим это утверждение. Рассмотрим множество V, состоящее из всех векторов ξ, для которых $\|\xi\| \leqslant 1$. Покажем, что это множество является выпуклым симметричным телом в E_n.

Выпуклость множества V проверяется так. Пусть $\xi \in V$ и $\eta \in V$, и пусть $0 \leqslant a \leqslant 1$. Тогда по свойствам нормы 3 и 4 $\|a\xi + (1-a)\eta\| \leqslant \|a\xi\| + \|(1-a)\eta\| = a\|\xi\| + (1-a)\|\eta\|$. Так как $\|\xi\| \leqslant 1$ и $\|\eta\| \leqslant 1$, то

$$\|a\xi + (1-a)\eta\| \leqslant a + (1-a) = 1,$$

т. е. вектор $a\xi + (1-a)\eta$ также принадлежит множеству V, что и означает выпуклость последнего.

Проверим, что O является внутренней точкой множества V. Действительно, если ξ любой вектор, не равный нулю, то, выбрав число $a = \frac{1}{\|\xi\|}$, мы убедимся, что

$$\|a\xi\| = a\|\xi\| = \frac{1}{\|\xi\|}\|\xi\| = 1,$$

*) В честь известного математика Г. Минковского — одного из создателей теории относительности.

т. е. $a\xi \in V$. Если же $\xi = \mathbf{0}$, то для любого числа a, $a\xi \in V$, так как по свойству 2 $\| a\xi \| = \| \mathbf{0} \| = 0 \leqslant 1$.

Симметричность множества V следует из свойства 3. Действительно, если $\xi \in V$, то $\| \xi \| \leqslant 1$, но тогда $\| -\xi \| = |-1| \| \xi \|$, т. е. $-\xi \in V$.

Ограниченность множества V доказывается несколько более громоздко, и мы это доказательство опустим.

Итак, множество V есть выпуклое тело и с помощью него можно определить норму $\| \xi \|_V$. Покажем, что эта норма совпадает с нормой $\| \xi \|$. Пусть a — такое положительное число, при котором $a\xi \in V$. Это значит, что $\| a\xi \| \leqslant 1$, а значит, $a \| \xi \| \leqslant 1$ и $\frac{1}{a} \geqslant \| \xi \|$. Нижняя грань величины $\frac{1}{a}$ достигается, если положить $a = \frac{1}{\| \xi \|}$, таким образом:

$$\inf_{a\xi \in V} \frac{1}{a} = \| \xi \|. \tag{55}$$

Сравнивая равенства (55) и (53), убеждаемся, что

$$\| \xi \| = \| \xi \|_V, \tag{56}$$

что и требовалось доказать.

В случае трехмерного пространства E_3 норме, определяемой с помощью произвольного выпуклого тела V, можно дать простое физическое истолкование.

Предположим, что имеется некоторая анизотропная среда, в которой звук распространяется с разной скоростью по разным направлениям. При этом скорость звука в любых двух противоположных направлениях одинакова.

Будем теперь из начала координат откладывать по каждому направлению вектор, равный величине скорости звука c вдоль этого направления. Предположим, что совокупность концов этих векторов ограничивает некоторое выпуклое тело V. Нетрудно видеть, что тело V ограничено, центрально-симметрично относительно начала координат и имеет последнее своей внутренней точкой. Тем самым в E_3 определяется норма $\| \xi \|_V$ и расстояние

$$\varrho_V(\xi, \eta) = \| \xi - \eta \|_V.$$

Читателю предоставляется проверить, что это расстояние равно времени, за которое звуковой сигнал проходит из конца вектора ξ в конец вектора η по прямой, соединяющей эти концы.

Кроме конечномерных пространств Минковского, в математике рассматриваются и аналогичные бесконечномерные — так называемые банаховы пространства. *Банаховым* *) *пространством* называется множество элементов (векторов), над которыми можно выполнять операции сложения, вычитания и умножения на число и для которых определено понятие нормы.

При этом нужно, чтобы операции над векторами удовлетворяли всем условиям, перечисленным на стр. 55, а норма обладала свойствами, указанными на стр. 62.

Пример бесконечномерного банахова пространства можно построить следующим образом.

Назовем вектором любую непрерывную функцию $f(t)$, определенную при $0 \leqslant t \leqslant 1$. Суммой векторов назовем сумму функций $f(t) + g(t)$. Произведением вектора $f(t)$ на число a назовем функцию $af(t)$, получаемую умножением всех ее значений на число a. Нулевым вектором является функция $f_0(t) \equiv 0$, тождественно равная нулю. Нормой функции назовем максимум ее модуля

$$\|f(t)\| = \max_{0 \leqslant t \leqslant 1} |f(t)|.$$

Читатель легко сумеет убедиться, что все свойства операций над векторами и все свойства нормы здесь выполнены.

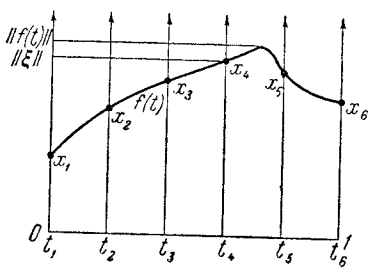

Рис. 28.

Это пространство обычно обозначается как $C_{[0,\,1]}$ или просто C.

Банаховы пространства, в которых векторами являются функции, играют большую роль в современной математике.

То обстоятельство, что метрическое пространство, точками которого являются функции, является бесконечномерным, можно пояснить следующим рассуждением.

Рассмотрим отрезок [0, 1] и проведем через какие-то из точек этого отрезка вертикальные прямые (рис. 28).

---

*) Название дано в честь известного польского математика С. Банаха (1892—1945), одного из основателей так называемого функционального анализа — важной ветви современной математики.

Возьмем вектор $\xi = (x_1, x_2, \ldots, x_n)$, принадлежащий n-мерному пространству, и будем откладывать его координаты на этих вертикальных прямых.

Полученные точки на вертикалях образуют график некоторой функции, заданной в соответствующих точках отрезка $[0, 1]$. Отсюда видно, что при $n \to \infty$ эти наборы точек могут приближаться к графикам непрерывных функций. (При этом надо выбирать только «гладкие» наборы точек, т. е. такие, что ординаты на близких вертикалях близки.) Если в E_n взять норму $C^{(n)}$, т. е. $\|\xi\| = \max(|x_1|, |x_2|, \ldots, |x_n|)$, то «в пределе» эта норма переходит в норму C:

$$\|f(t)\| = \max_{0 \leqslant t \leqslant 1} |f(t)|.$$

§ 8. СГЛАЖИВАНИЕ ОШИБОК ЭКСПЕРИМЕНТАЛЬНЫХ ИЗМЕРЕНИЙ

При наблюдении той или иной физической величины результатом эксперимента обычно является серия измеренных значений этой величины $(x_1, x_2, \ldots, x_n)$.

Сама эта величина может быть стационарной или переменной. В последнем случае значения $x_1, x_2, \ldots, x_n$ теоретически должны получаться изменяющимися по какому-то закону, а в первом случае постоянными. Кроме того, в величинах $x_1, x_2, \ldots, x_n$ содержатся еще ошибки измерений, искажающие картину.

Эту ситуацию можно было трактовать как получение некоего сообщения от природы, на которое из-за несовершенства эксперимента налагаются искажения.

Задача математической обработки измерений состоит в том, чтобы в какой-то степени восстановить правильное сообщение. К решению этой задачи можно применить соображения, уже использованные для автоматического исправления ошибок в дискретных сообщениях.

Если измеряемые величины могут принимать любые вещественные значения, то в качестве пространства сообщений можно взять n-мерное векторное пространство E_n. Расстояние между точками этого пространства $\varrho(\xi, \eta)$ определяется из конкретных свойств проводимого эксперимента. Наиболее часто используется расстояние вида

$$\varrho(\xi, \eta) = \sqrt{(x_1 - y_1)^2 + \ldots + (x_n - y_n)^2}, \qquad (57)$$

при котором пространство сообщений является пространством l_2^n.

Теоретически возможные сообщения образуют подмножество H пространства сообщений.

В качестве правильного сообщения выбирается вектор $\eta \in H$, «ближайший» к принятому сообщению ξ, т. е. такой, что

$$\varrho(\xi, \eta) = \min. \tag{58}$$

Для расстояния вида (57) этот принцип широко известен под названием метода наименьших квадратов и впервые предложен великим немецким математиком К. Гауссом.

Рассмотрим теперь конкретный пример подмножества теоретически возможных сообщений. Предположим, что измеряемая величина изменяется во времени по линейному закону:

$$y = kt + b\ ^*).$$

Это значит, что каждый вектор $\eta \in H$ имеет вид $\eta = (y_1, y_2, \ldots, y_n)$, где

$$\begin{aligned} y_1 &= kt_1 + b, \\ y_2 &= kt_2 + b, \\ &\ldots\ldots \\ y_n &= kt_n + b. \end{aligned}$$

Вектор $\eta \in H$ определяется двумя параметрами k и b.

Пусть вектор, полученный в результате измерений этой величины, равен $\xi = (x_1, x_2, \ldots, x_n)$. Основное условие (58) записывается теперь так:

$$F(k, b) = (kt_1 + b - x_1)^2 + (kt_2 + b - x_2)^2 + \ldots + \\ + (kt_n + b - x_n)^2 = \min.$$

В выражении $F(k, b)$ неизвестными являются параметры искомого теоретически возможного сообщения k и b; величины $t_1, t_2, \ldots, t_n$ и $x_1, x_2, \ldots, x_n$ известны из опыта.

---

*) Для простоты будем считать, что ошибками в определении моментов времени t_i можно пренебречь.

Для отыскания минимума величины $F(k, b)$ используем критерий, известный из дифференциального исчисления:

$$\frac{\partial F}{\partial k}=0; \quad \frac{\partial F}{\partial b}=0, \tag{59}$$

который в данном случае положительной квадратичной функции $F(k, b)$ является необходимым и достаточным для минимума.

Вычислим частные производные:

$$\left.\begin{aligned}\frac{\partial F}{\partial k}&=2t_1(kt_1+b-x_1)+\ldots+2t_n(kt_n+b-x_n),\\ \frac{\partial F}{\partial b}&=2\ (kt_1+b-x_1)+\ldots+2\ (kt_n+b-x_n).\end{aligned}\right\} \tag{60}$$

Обозначим для удобства

$$\begin{aligned}[t^2]&=t_1^2+t_2^2+\ldots+t_n^2,\\ [t]&=t_1+t_2+\ldots+t_n,\\ [tx]&=t_1x_1+t_2x_2+\ldots+t_nx_n,\\ [x]&=x_1+x_2+\ldots+x_n,\\ [1]&=1+1+\ldots+1=n.\end{aligned}$$

Выражения (60) можно записать тогда в виде

$$\left.\begin{aligned}\frac{\partial F}{\partial k}&=2[t^2]k+2[t]b-2[tx],\\ \frac{\partial F}{\partial b}&=2[t]k+2[1]b-2[x].\end{aligned}\right\} \tag{61}$$

Приравнивая согласно (59) эти выражения нулю, сокращая на два и перенося свободные члены в правую часть, получаем основные уравнения метода наименьших квадратов в символической форме:

$$\left.\begin{aligned}[t^2]k+[t]b&=[tx],\\ [t]k+[1]b&=[x].\end{aligned}\right\} \tag{62}$$

На рис. 29 изображены измерения значения x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8 и определенная по методу наименьших квадратов прямая $y=kt+b$. Из этого рисунка понятен смысл термина «сглаживание» ошибок.

Следующая таблица показывает рациональный порядок выполнения вычислений.

№ п. п.	x_i	t_i	$t_i x_i$	t_i^2
1	0,20	0,30	0,06	0,09
2	0,43	0,91	0,39	0,83
3	0,35	1,50	0,53	2,25
4	0,52	2,00	1,04	4,00
5	0,81	2,20	1,78	4,84
6	0,68	2,62	1,79	6,86
7	1,15	3,00	3,45	9,00
8	0,85	3,30	2,81	10,89
Σ	$[x]=4,79$	$[t]=15,83$	$[tx]=11,85$	$[t^2]=38,76$

Система (62) имеет в этом случае вид

$$\left.\begin{aligned} 38,76k + 15,83b &= 11,85, \\ 15,83k + 8b &= 4,79. \end{aligned}\right\}$$

Решением являются $k=0,319$, $b=-0,032$. Искомое теоретическое сообщение $y=0,319t-0,032$.

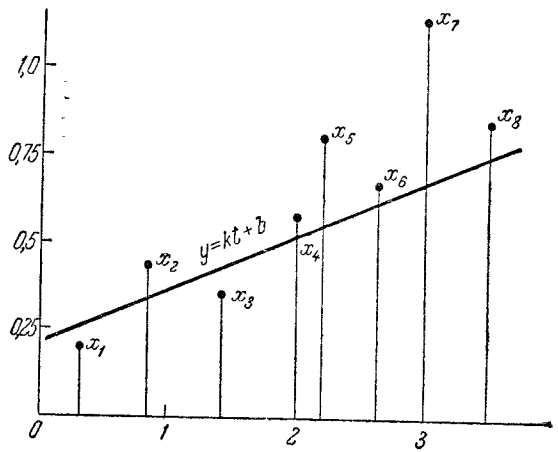

Рис. 29.

Аналогично, если множество теоретически возможных сообщений H состоит из параболических зависимостей

вида $y = at^2 + bt + c$, то основное условие (53) запишется в виде

$$F(a, b, c) = (at_1^2 + bt_1 + c - x_1)^2 + \ldots$$
$$\ldots + (at_n^2 + bt_n + c - x_n)^2 = \min.$$

Минимизация функции $F(a, b, c)$ сводится к решению системы трех линейных уравнений с тремя неизвестными параметрами a, b и c.

Основные приемы метода наименьших квадратов легко переносятся на случай расстояния вида

$$\varrho(\xi, \eta) = \alpha_1 (x_1 - y_1)^2 + \alpha_2 (x_2 - y_2)^2 + \ldots + \alpha_n (x_n - y_n)^2, \quad (63)$$

где α_1, α_2, ..., α_n — положительные числа (веса). Неравные веса приходится использовать, если отдельные измерения в эксперименте не равноточны и более грубым измерениям приходится приписывать наименьший вес.

Основной принцип сглаживания ошибок (58) применяется также для расстояний в пространствах $C^{(n)}$ и $l_1^{(n)}$. Однако при этом методы восстановления теоретически возможного сообщения оказались более сложными.

§ 9. БОЛЕЕ ОБЩИЕ ОПРЕДЕЛЕНИЯ РАССТОЯНИЯ

Как мы уже говорили, возможны различные обобщения понятия расстояния. Одно из наиболее радикальных обобщений используется в теории относительности. В этой теории рассматривается пространственно-временной мир, состоящий из точек вида (x, y, z, t), где x, y и z — пространственные координаты, а t — момент времени. Расстояние (пространственно-временной интервал) между двумя такими точками определяется по формуле

$$\varrho(\xi, \eta) = \sqrt{c^2 (t - t_1)^2 - (x - x_1)^2 - (y - y_1)^2 - (z - z_1)^2}, \quad (64)$$

где c — скорость света. Ясно, что расстояние (64) может принимать и мнимые значения.

Можно было бы обобщить понятие расстояния, считая, что функция $\varrho(M, N)$, удовлетворяя всем свойствам расстояния 1, 2, 3 и 4 (см. § 3, стр. 18), может принимать бесконечные значения *). Однако в этом случае множество

*) Точнее, значение $+\infty$.

точек пространства может быть разбито на отдельные подмножества без общих точек, каждое из которых является уже метрическим пространством в обычном смысле. Следовательно, такое обобщение не слишком интересно. Идея доказательства этого факта состоит в следующем.

Рассмотрим такое «обобщенное» пространство E. Выберем некоторую точку $N_1 \in E$ и рассмотрим множество E_1, состоящее из всех точек M пространства E, для которых расстояние $\varrho(N_1, M)$ конечно.

Пусть M и M' — произвольные точки множества E_1, тогда по аксиоме треугольника

$$\varrho(M, M') \leqslant \varrho(M, N_1) + \varrho(N_1, M').$$

Следовательно, $\varrho(M, M')$ конечно, а, значит, множество E_1 образует метрическое пространство с конечным расстоянием. Выкинем из E множество E_1. Если E совпадает с E_1, то само E есть обычное метрическое пространство. Если же нет, то среди оставшихся элементов выберем наугад элемент N_2 и образуем множество E_2 из элементов M, для которых расстояние $\varrho(N_2, M)$ конечно.

Продолжая этот процесс *), можно как раз и получить разбиение обобщенного пространства E на обычные метрические пространства с конечным расстоянием. Этот пример показывает, что получить содержательное обобщение какого-либо математического понятия не так-то просто. Во всяком случае, такое обобщение должно идти от содержательного изучения объектов, а не от чисто формального изменения аксиом. В частности, существует ряд вполне содержательных обобщений метрического пространства. Одно из таких обобщений мы и рассматриваем далее. Именно, понятие расстояния можно обобщить, отказавшись от условия симметрии (условия 1). Получающийся при этом класс пространств оказывается связанным с некоторыми существенными для математики объектами и поэтому заслуживает рассмотрения.

Мы будем называть *обобщенным метрическим пространством* множество E, для любых двух элементов которого ξ и η определено вещественное число $\varrho(\xi, \eta)$ со следующими свойствами:

1. $\varrho(\xi, \eta) \geqslant 0$.

*) В действительности проводить этот процесс придется с помощью так называемой трансфинитной индукции, поэтому мы говорим только об идее доказательства.

2. Двойное равенство $\varrho(\xi, \eta) = \varrho(\eta, \xi) = 0$ выполняется тогда и только тогда, когда элементы ξ и η совпадают:

$$\xi = \eta.$$

3. Для любой тройки элементов ξ, η, ζ

$$\varrho(\xi, \eta) \leqslant \varrho(\xi, \zeta) + \varrho(\zeta, \eta).$$

Ясно, что обычное расстояние удовлетворяет этим условиям. Однако этим свойствам 1, 2, 3 может удовлетворять и функция $\varrho(\xi, \eta)$, несимметричная относительно своих аргументов. Пример с несимметричным расстоянием у нас встретился фактически в конце § 4 в связи с определением расстояния как минимального времени, затрачиваемого на переход из одной точки в другую. В связи с тем, что путь в противоположные концы может занимать разное время, это расстояние, вообще говоря, несимметрично, хотя аксиома треугольника проверяется легко.

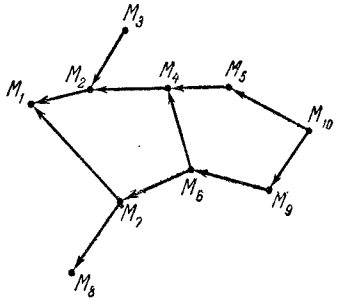

Рис. 30.

Другой пример пространства с несимметричным расстоянием получается, если рассмотреть множество из 10 точек, являющихся вершинами ломаной, изображенной на рис. 30.

Расстояние между точками M_i и M_j, $\varrho(M_i, M_j)$ определяется как минимальное число отрезков, проходимых против стрелки в пути, соединяющем M_i и M_j.

Например,

$$\varrho(M_1, M_{10}) = 4; \quad \varrho(M_{10}, M_1) = 0;$$
$$\varrho(M_3, M_9) = 3; \quad \varrho(M_9, M_3) = 1 \text{ и т. д.}$$

Ясно, что $\varrho(M_i, M_j) \geqslant 0$. Условие $\varrho(M_i, M_j) = \varrho(M_j, M_i) = 0$ (равенство нулю обоих расстояний) означает, что и точку M_i с M_j и точку M_j с M_i можно соединить отрезками, направленными по стрелке, т. е. M_i и M_j принадлежат замкнутой ломаной с согласованными направлениями отрезков. Так как таких петель на рис. 30 нет, то равенство $\varrho(M_i, M_j) = \varrho(M_j, M_i) = 0$ влечет за собой совпадение точек M_i и M_j. Таким образом, мы проверили выполнение условий 1 и 2 для несимметричного расстояния.

Аксиома треугольника проверяется следующим рассуждением. Рассмотрим путь с наименьшим количеством отрезков, направленных против стрелки, идущий из M_i в M_k, и аналогичный путь из M_k в M_j. Объединяя эти пути, мы получим путь из M_i в M_j с количеством отрезков, направленных против стрелки, равным $\varrho(M_i, M_k)+\varrho(M_k, M_j)$. Так как в «кратчайшем» пути из M_i в M_j количество таких отрезков может быть только меньше, то

$$\varrho(M_i, M_j) \leqslant \varrho(M_i, M_k)+\varrho(M_k, M_j). \tag{65}$$

В этом примере можно было бы определить новое расстояние по правилу

$$\varrho^*(M_i, M_j)=\varrho(M_i, M_j)+\varrho(M_j, M_i). \tag{66}$$

Это расстояние равно просто минимальному количеству отрезков в пути, соединяющем точки M_i и M_j. Отсюда ясно, что $\varrho^*(M_i, M_j)$ обладает свойствами обычного расстояния.

Аналогичный факт справедлив для любого обобщенного метрического пространства.

Теорема. *Если в обобщенном метрическом пространстве с расстоянием* $\varrho(\xi, \eta)$ *ввести новое расстояние*

$$\varrho^*(\xi, \eta)=\varrho(\xi, \eta)+\varrho(\eta, \xi), \tag{67}$$

то оно обладает всеми свойствами обычного расстояния и пространство превращается в обычное метрическое пространство.

Доказательство. Симметричность расстояния $\varrho^*(\xi, \eta)$ следует из того, что правая часть (67) не меняется при перестановке ξ и η. Так как $\varrho(\xi, \eta) \geqslant 0$ и $\varrho(\eta, \xi) \geqslant 0$, то $\varrho^*(\xi, \eta) \geqslant 0$. Равенство $\varrho^*(\xi, \eta)=0$ равносильно двойному равенству $\varrho(\xi, \eta)=\varrho(\eta, \xi)=0$, т. е. равносильно совпадению точек ξ и η. Наконец, складывая два неравенства

$$\varrho(\xi, \eta) \leqslant \varrho(\xi, \zeta)+\varrho(\zeta, \eta)$$

и

$$\varrho(\eta, \xi) \leqslant \varrho(\eta, \zeta)+\varrho(\zeta, \xi),$$

получаем

$$\varrho(\xi, \eta)+\varrho(\eta, \xi) \leqslant \varrho(\xi, \zeta)+\varrho(\zeta, \xi)+\varrho(\zeta, \eta)+\varrho(\eta, \zeta)$$

или

$$\varrho^*(\xi, \eta) \leqslant \varrho^*(\xi, \zeta)+\varrho^*(\zeta, \eta),$$

т. е. для нового расстояния выполнено неравенство треугольника.

Интересный пример обобщенного метрического пространства с несимметричным расстоянием получается с помощью важного понятия частично упорядоченного множества.

Множество называется *частично упорядоченным*, если для некоторых пар точек M и N этого множества имеет смысл отношение $M \prec N$ (читается: M предшествует N) и при этом выполняются свойства:

1. Если $M \prec N$, то невозможно отношение $N \prec M$.
2. Если $M \prec N$, а $N \prec L$, то $M \prec L$.

Примером частично упорядоченного множества является множество точек на рис. 30. Для этого нужно положить $M_j \prec M_i$, если из точки M_i можно достичь точку M_j, двигаясь только по отрезкам вдоль стрелки. Например, $M_8 \prec M_{10}$, $M_1 \prec M_3$, $M_2 \prec M_6$.

Второй пример получится, если взять множество точек на числовой прямой и положить $x \prec y$, если число $x < y$. В этом случае для каждой пары различных точек x и y справедливо либо отношение $x \prec y$, либо $y \prec x$ (такие множества называются упорядоченными).

Во всяком частично упорядоченном множестве E можно ввести понятие *непосредственного предшествования*.

Точка M непосредственно предшествует точке N (пишется: $M \bigcirc N$), если $M \prec N$ и не существует третьей, отличной от M и N точки L, лежащей между ними: $M \prec L \prec N$.

Так, в примере на рис. 30 $M_2 \bigcirc M_3$, $M_2 \bigcirc M_4$, $M_1 \bigcirc M_7$ и т. д.

Во множестве вещественных чисел непосредственно предшествующих точек нет вообще. Действительно, если $x \prec y$, то всегда $x \prec \frac{x+y}{2} \prec y$, так как $x < \frac{x+y}{2} < y$.

Теперь рассмотрим конечное частично упорядоченное множество E. Предположим, что множество E обладает свойством *связности*. Это значит, что для любых точек M и N из E существует последовательность точек $L_1 = M, L_2, \ldots, L_k = N$, так что для любой пары L_i, L_{i+1} либо $L_i \prec L_{i+1}$, либо $L_{i+1} \prec L_i$. Так, для точек M_3 и M_8 на рис. 30 эту последовательность можно составить так:

$$L_1 = M_3; \quad L_2 = M_2; \quad L_3 = M_7; \quad L_4 = M_8,$$

так как

$$M_2 \prec M_3; \quad M_1 \prec M_2; \quad M_8 \prec M_7.$$

Можно проверить, что это множество является связным.

Множество точек на рис. 31 не является связным, так как для точек M_8 и M_3 такой последовательности нельзя найти. Однако оба подмножества $E_1\{M_1, M_2, M_3, M_4\}$ и $E_2\{M_5, M_6, M_7, M_8\}$ являются связными. Можно легко проверить, что всякое конечное частично упорядоченное множество распадается на связные подмножества без общих точек. Введем теперь в конечное связное частично упорядоченное множество E расстояние $\varrho(M, N)$, определенное по следующему правилу. Назовем путем из M в N цепочку из точек $L_1 = M$; L_2, L_3, ..., $L_k = N$ таких, что всегда или $L_i \bigcirc L_{i+1}$, или $L_{i+1} \bigcirc L_i$. Длиной пути мы назовем количество точек L_i в нем, для которых $L_{i+1} \bigcirc L_i$. Тогда расстоянием $\varrho(M, N)$ мы назовем минимальную длину пути из M в N.

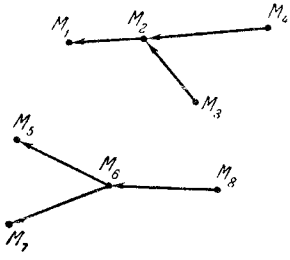

Рис. 31.

Проверим, что так определенное расстояние обладает всеми необходимыми свойствами.

Неотрицательность расстояния $\varrho(M, N)$ следует из определения. Для доказательства второго свойства расстояния заметим, что при несовпадающих M и N из условия $\varrho(M, N) = 0$ следует, что $M \{ N$. Действительно, условие $\varrho(M, N) = 0$ означает, что существует цепочка точек $L_1 = M$, L_2, ..., $L_k = N$ такая, что всегда $L_i \bigcirc L_{i+1}$, а следовательно, и $L_i \{ L_{i+1}$. Но тогда по второму свойству частичной упорядоченности получается, что $M \{ N$. Аналогично, из $\varrho(N, M) = 0$ вытекает, что $N \{ M$. Значит, если M и N — разные точки и $\varrho(M, N) = \varrho(N, M) = 0$, то одновременно $M \{ N$ и $N \{ M$. Так как последнее невозможно по первому свойству частично упорядоченных множеств, то M совпадает с N. Итак, второе свойство расстояния доказано.

Для доказательства третьего свойства (аксиомы треугольника) используем надежно послуживший нам ранее прием. Возьмем «кратчайший» путь из M в Q: $M = L_1, L_2 \ldots$ $\ldots, L_k = Q$ и «кратчайший» путь из Q в N: $Q = L_1', L_2', \ldots$ $\ldots, L_p = N$. Объединяя их, мы получим цепочку точек $M = L_1, L_2, \ldots, L_k = Q = L_1', L_2', \ldots, L_p' = N$, образующую путь из M в N. Общее число смежных точек этой цепочки, для которых $L_{i+1} \bigcirc L_i$ ($L_{i+1}' \bigcirc L_i'$), равно сумме

расстояний $\varrho(M, Q)+\varrho(Q, N)$. Ясно, что число таких пар для «кратчайшего» пути из M в N может быть только меньше этого значения:

$$\varrho(M, N) \leqslant \varrho(M, Q)+\varrho(Q, N). \tag{68}$$

Итак, доказано, что всякое конечное связное частично упорядоченное множество E является обобщенным метрическим пространством, если определить расстояние по указанному рецепту.

В качестве упражнения предоставляем читателю убедиться, что для всякой пары точек M и N таких, что $M \ll N$, расстояние $\varrho(M, N)=0$.

Оказывается, что верно и в некотором смысле обратное утверждение. Именно в каждом обобщенном метрическом пространстве E можно ввести частичную упорядоченность, определив правило предшествования так: $M \ll N$, если $\varrho(M, N)=0$.

Покажем, что при этом выполнены оба свойства частичной упорядоченности. Пусть M и N—разные точки и $M \ll N$, т. е. $\varrho(M, N)=0$. Тогда, если бы было одновременно $N \ll M$, то было бы и $\varrho(N, M)=0$. Отсюда по второму свойству расстояния вытекало бы, что M и N совпадают. Так как последнее неверно по предположению, то, значит, из $M \ll N$ следует невозможность отношения $N \ll M$. Первое свойство частичной упорядоченности доказано.

Пусть теперь $M \ll L$ и $L \ll N$. Это значит, что $\varrho(M, L)=0$ и $\varrho(L, N)=0$. Тогда, по аксиоме треугольника,

$$\varrho(M, N) \leqslant \varrho(M, L)+\varrho(L, N)=0,$$

т. е. $\varrho(M, N)=0$ и $M \ll N$.

Мы доказали, таким образом, что из $M \ll L$ и $L \ll N$ следует $M \ll N$.

Полученное частично упорядоченное множество, вообще говоря, не обязано быть связным.

Рассмотрим, например, частично упорядоченное множество E точек на рис. 31 (стр. 74). Между точками подмножества $E_1\{M_1, M_2, M_3, M_4\}$ можно определить расстояние с помощью «наикратчайших» путей. То же самое можно сделать и для точек подмножества $E_2\{M_5, M_6, M_7, M_8\}$. Определим теперь расстояние между любой точкой $M_i \in E_1$ и любой точкой $M_j \in E_2$, просто положив

$$\varrho(M_i, M_j)=\varrho(M_j, M_i)=100. \tag{69}$$

Нетрудно проверить, что мы получим обобщенное метрическое пространство, в котором равенство $\varrho(M, N)=0$ равносильно отношению $M \parallel N$ в частично упорядоченном множестве E. Однако E не является, как уже говорилось, связным частично упорядоченным множеством.

Можно ввести понятие *связного обобщенного метрического пространства*. Именно, пространство E называется *связным*, если для любой пары несовпадающих точек M и N из E существует цепочка точек $L_1=M, L_2, \ldots, L_k=N$ такая, что для любой смежной пары точек L_i и L_{i+1}, либо $\varrho(L_i, L_{i+1})=0$, либо $\varrho(L_{i+1}, L_i)=0$. Читатель легко проверит, что связным метрическим пространствам соответствуют связные частично упорядоченные множества.

Конечное частично упорядоченное множество (а значит, и соответствующее метрическое пространство) можно довольно просто изобразить геометрически. Для этого будем изображать элементы частично упорядоченного множества точками в трехмерном пространстве, обозначая их теми же буквами, что и соответствующие элементы. Каждую пару точек M и N, для которых $M \bigcirc N$, соединим линией, направленной из N в M, и укажем это направление стрелкой. Полученная геометрическая фигура, состоящая из точек (вершин этой фигуры) и соединяющих их направленных линий, называется *графом*. Примеры графов мы уже видели на рис. 30 и 31.

Нетрудно видеть, что если $M \parallel N$, то из N в M можно попасть, двигаясь только по направлению стрелки.

Заметим в заключение, что пространства с несимметричным расстоянием связаны также с важным для математики понятием дискретных топологических пространств.

На этом мы заканчиваем изучение понятия расстояния. Мы установили, что это понятие имеет много различных аспектов и связано не только с чисто математическими задачами, но и с такими вполне прикладными проблемами, как построение помехоустойчивых кодов. Эта особенность — многозначность применений и сложные логические связи — характерна и для других существенных математических понятий. Основной смысл создания таких понятий как раз и состоит в возможности установления связей и аналогий между, казалось бы, очень далекими областями и в выяснении скрытых пружин, от которых зависят математические свойства различных объектов.